3.6 Asymptotic normality of M-estimates. First, let's note some of the conditions under which nonlinear functions of sample averages are asymptotically normal. Let f be a function from an open interval containing a point μ into $\mathbb{R}$. Suppose the derivative $f'(\mu)$ exists and is not 0. Let $X_1, X_2, \ldots,$ be i.i.d. variables with mean μ and variance $0 < \sigma^2 := \sigma^2(X_1) < \infty$. Let $S_n := X_1 + \cdots + X_n$, and $\overline{X}_n := S_n/n$. Then $|\overline{X}_n - \mu| = O_p(n^{-1/2})$ by Chebyshev's inequality, and

$$f(\overline{X}_n) = f(\mu) + f'(\mu)(\overline{X}_n - \mu) + o(|\overline{X}_n - \mu|), \text{ so}$$
$$n^{1/2}(f(\overline{X}_n) - f(\mu)) = f'(\mu)((S_n - n\mu)/n^{1/2}) + o_p(1).$$

Thus the distribution of the left side converges to $N(0, f'(\mu)^2\sigma^2)$ by the central limit theorem. This kind of reasoning is known as the "delta-method." To extend the method to vector-valued random variables, let f be a real-valued function on an open set $U \subset \mathbb{R}^k$. Then f is said to be *Fréchet differentiable* at a point $t \in U$ if there is a vector $v := f'(t) \in \mathbb{R}^k$ such that

$$f(u) = f(t) + v \cdot (u - t) + o(|u - t|)$$

as $u \to t$. For $k = 1$, this is equivalent to the usual derivative. For $k > 1$, the components of $f'(t)$ will be the partial derivatives $\partial f(u)/\partial u_i|_{u=t}$, forming the gradient of f at t. Each partial derivative is a directional derivative in the direction of a coordinate axis. Existence of the Fréchet derivative means that not only these partial derivatives exist, but the graph of f has a tangent hyperplane at $(t, f(t)) \in \mathbb{R}^{k+1}$, see Problem 2.

Using the central limit theorem in $\mathbb{R}^k$, the delta-method extends straightforwardly to $\mathbb{R}^k$-valued random variables having finite second moments.

If f takes values in $\mathbb{R}^m$ then the definition of Fréchet derivative is formally the same, but with the vector v replaced by a linear transformation from $\mathbb{R}^k$ into $\mathbb{R}^m$, given by an $m \times k$ matrix. The Fréchet differentiability of f is equivalent to that of each of its m component real-valued functions.

Now, let's consider an exponential family of order m in a minimal representation (2.5.3) with densities $e^{\theta \cdot T(x) - j(\theta)}$, defined for $\theta \in \Theta \subset \mathbb{R}^m$, the interior of the natural parameter space. Recall Theorem 3.3.17, on MLEs for exponential families, and its proof. Let $\overline{T}_n := \frac{1}{n}\sum_{i=1}^n T(X_i)$. The maximum likelihood estimator of θ, when it exists, is $(\nabla j)^{-1}(\overline{T}_n)$. Here ∇j has a continuous derivative which is an $m \times m$ matrix whose entries are second partial derivatives of j. The matrix is nonsingular and so has a nonsingular inverse, which is the derivative of $(\nabla j)^{-1}$ (e.g. Rudin, 1976, p. 223, (52)). Since $\overline{T}_n$ is asymptotically normal by the multidimensional central limit theorem (RAP, Theorem 9.5.6), it follows by the delta-method that the MLE is also.

Next, asymptotic normality will be shown in quite general cases. Let Θ be an open subset of $\mathbb{R}^m$, $(X, \mathcal{A}, P)$ a probability space, and ψ a function from $\Theta \times X$ into $\mathbb{R}^m$. Let $X_1, X_2, \ldots$ be i.i.d. with values in X and distribution P. It will be shown under further assumptions that for a sequence $T_n = T_n(X_1, \ldots, X_n)$ of statistics with values in Θ, if

(3.6.1) $$n^{-1/2}\sum_{i=1}^n \psi(T_n, X_i) \to 0 \text{ in probability,}$$

then the distribution of T_n will converge to some normal law.

Recall that the definition (3.5.1) of approximate M-estimators of ψ type had a further factor of $n^{-1/2}$ as compared to (3.6.1), although it also assumes a stronger form of convergence. So it seems a relatively mild condition to assume that the sequence $\{T_n\}$ is consistent, as is shown under different conditions in Secs. 3.3 and 3.5. Also, asymptotic normality with mean 0 of $n^{1/2}(T_n - \theta_0)$ implies consistency in probability. At any rate, it will be an assumption here:

(AN-1) For some θ_0, $T_n(X_1, \ldots, X_n) \rightarrow \theta_0$ in probability as $n \to \infty$.

In the next assumption, that each $\psi(\theta, \cdot)$ is measurable is the same as (B-1) in Sec. 3.5:

(AN-2) For each $\theta \in \Theta$, $\psi(\theta, \cdot)$ is measurable, and $\psi(\cdot, \cdot)$ is separable.

The next condition is similar to (B-3) in Sec. 3.5:

(AN-3) $\lambda(\theta) := E\psi(\theta, x)$ is defined and finite for all θ, and for θ_0 in (AN-1), $\lambda(\theta_0) = 0$.

The further condition in (B-3), that $\lambda(\theta) \neq 0$ for $\theta \neq \theta_0$, is not needed here because of (AN-1), and for θ near θ_0, (AN-4)(i) below.

The norm $|\cdot|$ on $\mathbb{R}^m$ will be taken to be $|\theta| := \max(|\theta_1|, \ldots, |\theta_m|)$. Then if $\|\cdot\|$ is the usual Euclidean norm, we have $|\theta| \leq \|\theta\| \leq m^{1/2}|\theta|$ for all θ, so (like any two norms on $\mathbb{R}^m$) the two norms are equivalent within constant multiples.

For the next condition, given any $\theta \in \Theta$, since Θ is open, for $\delta > 0$ small enough, $|\phi - \theta| < \delta$ implies $\phi \in \Theta$. For such a $\delta > 0$ depending on θ, let

$$u(\theta, x, \delta) := \sup\{|\psi(\phi, x) - \psi(\theta, x)| : |\phi - \theta| \leq \delta\}.$$

(This is called the modulus of continuity of $\psi(\cdot, x)$ at θ.) The next assumption is:

(AN-4) For some numbers $a > 0$, $b > 0$ and $\gamma > 0$,
(i) for $|\theta - \theta_0| \leq \gamma$, we have $\theta \in \Theta$ and $|\lambda(\theta)| \geq a|\theta - \theta_0|$;
(ii) $\max(Eu(\theta, x, t), E[u(\theta, x, t)^2]) \leq bt$ for any $t \geq 0$ such that $|\theta - \theta_0| \leq \gamma - t$.

The last assumption is

(AN-5) $E|\psi(\theta_0, x)|^2 < \infty$.

In practice, since θ_0 is unknown, we would need to have a function $\psi(\cdot, \cdot)$ such that $E|\psi(\theta, x)|^2 < \infty$ for all θ. Since P is also unknown, this means either that we assume P belongs to a family of laws for which the given assumptions hold, or we choose a ψ such that they hold for all P. For (AN-5) this would mean that $\psi(\theta, \cdot)$ is a bounded function of x for each θ.

On first reading, I advise skipping to Lemma 3.6.13 and its "heuristic proof," then to the main Theorem 3.6.15 on asymptotic normality and its short proof.

The detailed and rigorous proof (by Huber) is as follows. For any τ and θ let $\eta(\theta, x) := \psi(\theta, x) - \lambda(\theta)$ and

$$Z_n(\tau, \theta) := \left|\textstyle\sum_{i=1}^n [\eta(\tau, X_i) - \eta(\theta, X_i)]\right| / (n^{1/2} + n|\lambda(\tau)|).$$

Most of the work in proving asymptotic normality will be in the following:

3.6.2 Lemma. If assumptions (AN-2), (AN-3) and (AN-4) hold for some θ_0, then

$$\sup\{Z_n(\tau,\theta_0):\ |\tau-\theta_0|\leq\gamma\}\ \rightarrow\ 0$$

in probability as $n\to\infty$.

Proof. By an affine change of coordinates, take $\theta_0=0$ and $\gamma=1$ (this may change the values of $a,\ b$).

The cube $|\tau|\leq 1$ will be decomposed into smaller cubes on which $Z_n(\tau,0)$ will be bounded separately. Recall that $\lceil x\rceil$ is the smallest integer $\geq x$. Given $\varepsilon>0$, let

$$M\ :=\ \lceil\max(2,(3b)/(a\varepsilon))\rceil. \tag{3.6.3}$$

Let $q\ :=\ 1/M$. Let K be a positive integer and consider the cubes centered at the origin, $C_k\ :=\ \{\theta:\ |\theta|\leq(1-q)^k\}$, for $k=0,1,\ldots,K$. For $k\geq 1$, decompose the difference $C_{k-1}\setminus C_k$ into smaller cubes, as shown in Fig. 3.6A, with edges of length $\ell\ :=\ \ell_k\ :=\ (1-q)^{k-1}q$. Let $d\ :=\ \ell/2$. Then the centers of the smaller cubes are points ξ whose coordinates are odd multiples of d, with $|\xi|=(1-q)^{k-1}(1-q/2)$. (A reason for using the norm $|\cdot|$ is to make the latter norms $|\xi|$ all equal, as they would not be for other norms such as the Euclidean norm.)

For each $k\geq 1$, $C_{k-1}\setminus C_k$ is decomposed into less than $2m(2M)^{m-1}$ of the smaller cubes, so the grand total number N of the smaller cubes satisfies $N<2^mKmM^{m-1}$. Let them be numbered $B_1,\ldots,B_N$. K will be chosen depending on n, specifically $K=K(n)$ is the unique integer such that

$$(1-q)^K\ \leq\ n^{-3/4}\ <\ (1-q)^{K-1}, \tag{3.6.4}$$

or equivalently

$$K(n)-1<(3/4)\log(n)/|\log(1-q)|\ \leq\ K(n),$$

which implies

$$N=O(\log n)\ \text{ as }\ n\to\infty. \tag{3.6.5}$$

Now,

$$\Pr\{\sup\{Z_n(\tau,0):\ \tau\in C_0\}\geq 2\varepsilon\}\ \leq$$

$$\Pr(\sup\{Z_n(\tau,0):\ \tau\in C_K\}\geq 2\varepsilon)+\textstyle\sum_{j=1}^N\Pr(\sup\{Z_n(\tau,0):\ \tau\in B_j\}\geq 2\varepsilon). \tag{3.6.6}$$

It will be shown that the right side of (3.6.6) goes to 0 as $n\to\infty$, which will prove the Lemma.

Consider any of the cubes B_j with center ξ and edges of length $\ell=2d=(1-q)^{k-1}q$, for some $k=1,\ldots,K$. For any $\tau\in B_j$, we have by (AN-4) $|\lambda(\tau)|\ \geq\ a|\tau|\ \geq\ a(1-q)^k$, and so

$$|\lambda(\tau)-\lambda(\xi)|\ \leq\ Eu(\xi,x,d)\ \leq\ bd\ \leq\ b(1-q)^kq. \tag{3.6.7}$$

Next,

$$Z_n(\tau,0)\ \leq\ Z_n(\tau,\xi)+|\textstyle\sum_{i=1}^n[\eta(\xi,X_i)-\eta(0,X_i)]/(n^{1/2}+n|\lambda(\tau)|),\ \text{ so}$$

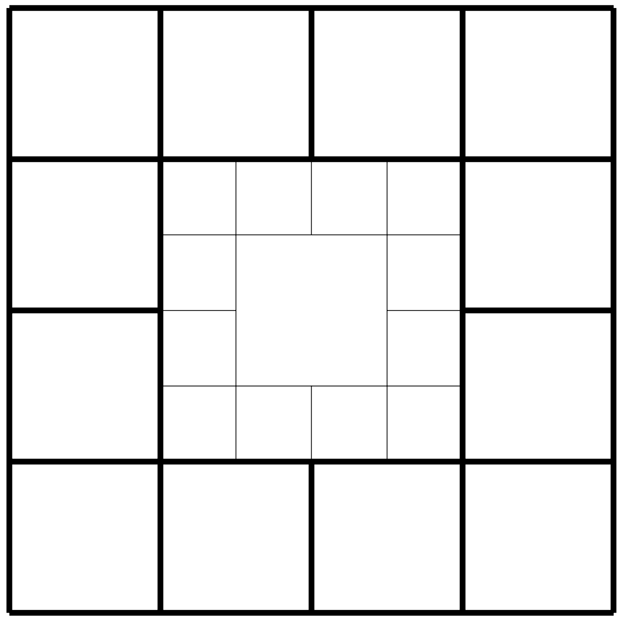

Figure 3.6A: $M = 2$

Fig. 3.6 A. Decomposition of cubes: proof of Lemma 3.6.2

$$\text{(3.6.8)} \qquad \sup\{Z_n(\tau,0):\ \tau\in B_j\} \ \le\ U_n+V_n$$

where

$$U_n \ :=\ \textstyle\sum_{i=1}^n[u(\xi,X_i,d)+Eu(\xi,x,d)]/(na(1-q)^k),$$
$$V_n \ :=\ \textstyle\sum_{i=1}^n[\eta(\xi,X_i)-\eta(0,X_i)]/(na(1-q)^k).$$

Then

$$\Pr(U_n\ge\varepsilon) \ =\ \Pr\{\textstyle\sum_{i=1}^n[u(\xi,X_i,d)-Eu(\xi,x,d)] \ \ge\ \varepsilon na(1-q)^k-2nEu(\xi,x,d)\}.$$

By (3.6.7) and (3.6.3),

$$\varepsilon a(1-q)^k-2Eu(\xi,x,d) \ \ge\ \varepsilon a(1-q)^k-2bq(1-q)^k \ \ge\ bq(1-q)^k,$$

which by (AN-4)(ii) and Chebyshev's inequality implies

$$\text{(3.6.9)} \qquad \Pr(U_n\ge\varepsilon) \ \le\ 1/[bqn(1-q)^k].$$

For V_n,

$$E[(\eta(\xi,\cdot)-\eta(0,\cdot))^2] \ =\ \mathrm{var}[\eta(\xi,\cdot)-\eta(0,\cdot)] \ =\ \mathrm{var}[\psi(\xi,\cdot)-\psi(0,\cdot)]$$
$$\le\ E[u(0,\cdot,|\xi|)^2] \ <\ b(1-q)^{k-1}.$$

Then by Chebyshev's inequality again, since (3.6.3) implies $b/(a^2\varepsilon^2)\le 1/(9bq^2)$, we have

$$\text{(3.6.10)} \qquad \Pr(V_n\ge\varepsilon) \ \le\ 1/[9nbq^2(1-q)^{k+1}].$$

From (3.6.4), (3.6.8), (3.6.9) and (3.6.10) we get

$$\text{(3.6.11)} \qquad \Pr(\sup\{Z_n(\tau,0):\ \tau\in B_j\}\ge 2\varepsilon) \ \le\ Jn^{-1/4}$$

where $J \ :=\ [bq(1-q)]^{-1}+[9bq^2(1-q)^2]^{-1}$. Thus by (3.6.5), as $n\to\infty$

$$\Pr(\sup\{Z_n(\tau,0):\ \tau\in C_0\setminus C_K\}\ge 2\varepsilon) \ \le\ NJn^{-1/4} \ =\ O(n^{-1/4}\log n).$$

Also,

$$\sup\{Z_n(\tau,0):\ \tau\in C_K\} \ \le\ n^{-1/2}\textstyle\sum_{i=1}^n[u(0,X_i,\delta)+Eu(0,x,\delta)]$$

for $\delta \ :=\ (1-q)^K \ \le\ n^{-3/4}$ by (3.6.4). So,

$$\Pr\left(\sup\{Z_n(\tau,0):\ \tau\in C_K\}\ge 2\varepsilon\right) \ \le$$
$$\Pr\left\{\textstyle\sum_{i=1}^n[u(0,X_i,\delta)-Eu(0,x,\delta)]\ge 2n^{1/2}\varepsilon-2nEu(0,x,\delta)\right\}.$$

Since by (AN-4)(ii) $Eu(0,x,\delta) \ \le\ b\delta \ \le\ bn^{-3/4}$, there is an n_0 such that

$$2n^{1/2}\varepsilon-2nEu(0,x,\delta) \ \ge\ n^{1/2}\varepsilon \ \text{ for } \ n\ge n_0.$$

Then by Chebyshev's inequality, since $\mathrm{var}(u(0,x,\delta))\le E[u(0,x,\delta)^2]\le b\delta\le bn^{-3/4}$,

$$\text{(3.6.12)} \qquad \Pr(\sup\{Z_n(\tau,0):\ \tau\in C_0\}\ge 2\varepsilon) \ \le\ O(n^{-3/4})+O(n^{-1/4}\log n),$$

which gives Lemma 3.6.2. □

The next step toward asymptotic normality is:

3.6.13 Lemma. Assume that (AN-1) through (AN-5) hold and T_n are estimators satisfying (3.6.1). For the rigorous proof, instead of (AN-1), it suffices to assume that $\Pr\{|T_n - \theta_0| \le \gamma\} \to 1$ as $n \to \infty$. Then

$$n^{1/2}\left(\lambda(T_n) + \frac{1}{n}\sum_{i=1}^n \psi(\theta_0, X_i)\right) \to 0 \text{ in probability.}$$

Notes. As $n \to \infty$, the second term being multiplied by $n^{1/2}$ converges to $\lambda(\theta_0)$, which is 0 by (AN-3). The second term is $O_p(n^{-1/2})$ but not $o_p(n^{-1/2})$ in general, by the central limit theorem. Thus the Lemma implies that $\lambda(T_n)$ is of the same order. See Problem 1.

(AN-1) states that $T_n \to \theta_0$ in probability, and so $\Pr\{|T_n - \theta_0| \le \gamma\} \to 1$. The proof will show that from (3.6.1), $T_n \to \theta_0$ at an $O_p(1/\sqrt{n})$ rate.

Heuristic proof of Lemma 3.6.13. Here Lemma 3.6.2 won't be applied. Assumptions (AN-4) about $E[u(\theta, x, t)^2]$ for $\theta = \theta_0$ and $t = \gamma$ and (AN-5) imply that for all θ in a neighborhood U of θ_0, namely $|\theta - \theta_0| < \gamma$, we have $E(|\psi(\theta, \cdot)|^2) < \infty$. So, we can apply the multidimensional central limit theorem (RAP, Theorem 9.5.6) to get that for $\theta \in U$, the distribution of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(\theta, X_i) - \psi(\theta_0, X_i) - \lambda(\theta)$$

converges as $n \to \infty$ to a normal distribution $N(0, C)$ where the covariance matrix $C = C^{\theta,\theta_0}$ depends on θ and θ_0. (Recall that $\lambda(\theta_0) = 0$.) As $\theta \to \theta_0$, $E(|\psi(\theta, \cdot) - \psi(\theta_0, \cdot)|^2) \to 0$ by (AN-4)(ii), and so $\mathrm{Var}(\psi_j(\theta, \cdot) - \psi_j(\theta_0, \cdot)) \to 0$ for $j = 1, \ldots, m$. It follows that the matrix $C^{\theta,\theta_0} \to 0$ as $\theta \to \theta_0$.

Now, substituting $\theta = T_n$ and recalling that as $n \to \infty$, $T_n \to \theta_0$ in probability by (AN-1), and $n^{-1/2}\sum_{i=1}^n \psi(T_n, X_i) \to 0$ in probability, we are left with $D_n := n^{-1/2}\sum_{i=1}^n(-\psi(\theta_0, X_i) - \lambda(T_n))$ having asymptotic distribution $N(0, C^{T_n,\theta_0})$, but since $T_n \to \theta_0$, $D_n \to 0$ in probability. With a minus sign, this gives the conclusion of Lemma 3.6.13.

Rigorous proof of Lemma 3.6.13. As in the proof of Lemma 3.6.2, assume that $\theta_0 = 0$ and $\gamma = 1$. Note that

$$\textstyle\sum_{i=1}^n \psi(T_n, X_i) \;=\; \sum_{i=1}^n[\psi(T_n, X_i) - \psi(0, X_i) - \lambda(T_n)] + \sum_{i=1}^n[\psi(0, X_i) + \lambda(T_n)].$$

So, with probability converging to 1,

$$\textstyle|\sum_{i=1}^n[\psi(0, X_i) + \lambda(T_n)]|/(n^{1/2} + n|\lambda(T_n)|) \;\le$$
$$\textstyle\sup\{Z_n(\tau, 0) :\ |\tau| \le 1\} \;+\; n^{-1/2}|\sum_{i=1}^n \psi(T_n, X_i)|.$$

The terms on the right approach 0 in probability as $n \to \infty$ by Lemma 3.6.2 and (3.6.1), so the left side does also.

Given $0 < \varepsilon < 1$, let $V^2 := 2E(|\psi(0, x)|^2)/\varepsilon$, which is finite by (AN-5), and $V \ge 0$. Then by Chebyshev's inequality, there is an n_0 such that for $n \ge n_0$, the inequalities

$$\textstyle n^{-1/2}|\sum_{i=1}^n \psi(0, X_i)| \;\le\; V$$

(3.6.14) and

$$|\sum_{i=1}^n[\psi(0,X_i)+\lambda(T_n)]| \ \le \ \varepsilon(n^{1/2}+n|\lambda(T_n)|)$$

each hold except with probability $\le \varepsilon/2$, so both are true on an event with probability at least $1-\varepsilon$. The latter inequality implies

$$n^{1/2}|\lambda(T_n)|(1-\varepsilon) \ \le \ \varepsilon + n^{-1/2}|\sum_{i=1}^n \psi(0,X_i)|,$$

so when (3.6.14) holds, $n^{1/2}|\lambda(T_n)| \ \le \ (V+\varepsilon)/(1-\varepsilon)$. So for $n \ge n_0$, with probability at least $1-\varepsilon$, we have

$$|n^{1/2}\lambda(T_n) + n^{-1/2}\sum_{i=1}^n \psi(0,X_i)| \ \le \ \varepsilon\left(1+\frac{\varepsilon+V}{1-\varepsilon}\right) \ = \ \varepsilon\cdot\frac{V+1}{1-\varepsilon}.$$

Letting $\varepsilon \downarrow 0$, $V = O(\varepsilon^{-1/2})$, $V\varepsilon = O(\varepsilon^{1/2})$, and the lemma follows. □

Since $\lambda(\cdot)$ takes the open set $\Theta \subset \mathbb{R}^m$ into $\mathbb{R}^m$, its Fréchet derivative at θ_0, if it exists, is a linear transformation A from $\mathbb{R}^m$ into itself, given by an $m \times m$ matrix. Note that if A exists and is non-singular, then (AN-4)(i) follows. Let B' denote the transpose of a matrix B.

3.6.15 Theorem. Assume (AN-1) through (AN-5), that T_n satisfy (3.6.1), and that λ has a non-singular Fréchet derivative A at θ_0. Then $n^{1/2}(T_n-\theta_0)$ is asymptotically normal with mean 0 and covariance matrix $A^{-1}C(A^{-1})'$, where C is the covariance matrix of $\psi(\theta_0,x)$.

Proof. Since $\lambda(\theta_0)=0$, we have $|\lambda(\theta)-A(\theta-\theta_0)| \ = \ o(|\theta-\theta_0|)$ as $|\theta-\theta_0| \to 0$.

By Lemma 3.6.13, the central limit theorem (RAP, Sec. 9.5), and Lemma 11.9.4 of RAP, as $n \to \infty$, $n^{1/2}\lambda(T_n)$ has distribution converging to $N(0,C)$ (a minus sign doesn't change this distribution). Thus $|A(T_n-\theta_0)| + o(|T_n-\theta_0|) = O_p(n^{-1/2})$. Since A is non-singular, $o(|T_n-\theta_0|) = o(|A(T_n-\theta_0)|)$. By (AN-1), $T_n \to \theta_0$ in probability. It follows then that $|A(T_n-\theta_0)| = O_p(n^{-1/2})$, and $|T_n-\theta_0| = O_p(n^{-1/2})$, so $o(|T_n-\theta_0|) = o_p(n^{-1/2})$, and $A(\sqrt{n}(T_n-\theta_0)) \to N(0,C)$ in distribution, again using (RAP, Lemma 11.9.4). Thus by the continuous mapping theorem (RAP, Theorem 9.3.7) applied to A^{-1},

$$\sqrt{n}(T_n-\theta_0) \ \to \ N(0,C)\circ(A^{-1})^{-1} \ = \ N(0,A^{-1}C(A^{-1})')$$

(RAP, Proposition 9.5.12). □

PROBLEMS

1. Let $\psi(\theta,x) \ := \ x-\theta$. Suppose the observations $X_1,X_2,\dots$ are i.i.d. for a probability measure P such that $\int x dP = \theta_0$. Suppose that in (3.6.1) the estimators T_n are chosen so that the given expression is 0 rather than just converging to 0. Then evaluate T_n and $\lambda(\cdot)$ explicitly. Show that the expression which converges to 0 in probability in the conclusion of Lemma 3.6.13 is actually identically 0 in this case.
2. Let (r,θ) be the usual polar coordinates on $\mathbb{R}^2$. Let $f(x,y) \ := \ \sin(2\theta)$ for $r>0$ and $f(0,0) \ := \ 0$. Show that the partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ exist at all points and are both 0 at $(0,0)$, but f is not Fréchet differentiable or even continuous at $(0,0)$.

3. Let $f(x) := x + x^2 \sin(1/x^2)$ for $x \neq 0$ and $f(0) := 0$. Show that f is differentiable at all x and $f'(0) = 1$, but f' is unbounded in every neighborhood of 0, so f is not C^1. Also, f is not one-to-one in any neighborhood of 0.

NOTES

This section is based on Sec. 4 of Huber (1967). If the function $\lambda(\cdot)$ is C^1, and has a non-singular derivative at θ_0, then a C^1 local inverse function λ^{-1} exists from a neighborhood of 0 to a neighborhood of θ_0, by the inverse function theorem, e.g. Rudin (1976, p. 223, (52)), or Hoffman (1975, Sec. 8.5, Theorem 7), and the delta-method applies to the inverse function. It turns out to be unnecessary to assume existence of the Fréchet derivative of λ except at θ_0 in Theorem 3.6.15 but, since θ_0 is unknown, in practice we would need the differentiability everywhere and in the cases actually encountered in statistics, it seems that $\lambda(\cdot)$ will then be C^1. Non-C^1 functions having derivatives everywhere exist even in one dimension (Problem 3) but seem to be rather pathological.

REFERENCES

Hoffman, Kenneth (1975). *Analysis in Euclidean Space.* Prentice-Hall, Englewood Cliffs, NJ.

Huber, P. J. (1967). See Sec. 3.3.

Rudin, Walter (1976). *Principles of Mathematical Analysis*, 3d ed. McGraw-Hill, New York.