The Annals of Statistics
1973, Vol. 1, No. 3, 557–561

ESTIMATION OF THE DERIVATIVE OF THE LOGARITHM OF A DENSITY[1]

BY VÁCLAV FABIAN

Michigan State University

Suppose g is a probability density on $R = (-\infty, +\infty)$ with a continuous derivative and with $I(g) = \int (g'/g)^2 g < +\infty$. Suppose $\{\varepsilon_n\}$ is a sequence of positive numbers converging to zero and V_i, Z_i are random variables, $Z_i \to 0$ with probability one and V_n is conditionally (given $V_1, \cdots, V_{n-1}, Z_1, \cdots, Z_n$) distributed according to g. Estimates h_n of g'/g are constructed, which are based on $V_1 + Z_1, \cdots, V_n + Z_n$ and have the following properties. For almost all ω, $h_n(\omega, \cdot) \to g'/g$ on $\{t; g(t) > 0\}$. If $n\varepsilon_n \uparrow +\infty$, $\sum n^{-1}\varepsilon_n^{-1}|Z_n| < +\infty$ a.e. then for almost all ω, $h_n(\omega, \cdot) \to g'/g$ in $L_2(g)$ and $\int h_n^2 \, dG_n \to I(g)$ where G_n is the empirical distribution function of $V_1 + Z_1, \cdots, V_n + Z_n$. The results on the pointwise and $L_2(g)$ convergences hold also if h_n are replaced by $h_n(\omega, v + \eta_n(v, \omega))$ provided η_n are small and preserve the measurability of the estimators.

1. Introduction. The result derived here is needed in another paper (Fabian (1973)) on asymptotically efficient stochastic approximation procedures. There are known results about estimation of $g'(G^{-1})/g(G^{-1})$ where G is the distribution function with density g (Hájek and Šidák (1967), van Eeden (1970), Weiss and Wolfowitz (1970)) and results about estimating g and its derivatives (cf. Schuster (1969) and references given there).

Estimators of g'/g could be obtained from the results in Schuster (1969) but under stronger conditions than are required here and the additional properties (data contaminated by the Z_i's, possibility of contaminating h_n by the η_n's, cf. Extension 2.3) would need an additional proof. It seemed then easier to construct the estimates and prove their properties *ab initio*. The proof has, however, much in common with the proof in Hájek and Šidák (1967), VII. 1.5.

2. The results. Assume $(\Omega, \mathscr{F}, P)$ is a probability space. Unless specified otherwise, convergence for random variables is meant with probability one. If h is a function on R, the real line, then $h^c(x)$ denotes $h(x + c) - h(x - c)$.

(2.1) LEMMA. *Suppose G is a distribution function, $V_1, V_2, \cdots, Z_1, Z_2, \cdots$ are random variables and the conditional distribution of V_n, given $V_1, \cdots, V_{n-1}, Z_1, \cdots, Z_n$, is given by G. Let G_n be the empirical distribution function of $V_1 + Z_1, \cdots, V_n + Z_n$ and $F_n(t) = n^{-1} \sum_{j=1}^n G(t - Z_j)$. Then for every $0 \leqq r < \frac{1}{2}$*

$$n^r \sup_t |G_n(t) - F_n(t)| \to 0 \,.$$

PROOF. Fix a t and set $\Delta_n = n(G_n(t) - F_n(t))$, $I_j = \chi_{\{V_j + Z_j \leqq t\}}$, $p_j = G(t - Z_j)$, $\mathscr{F}_j = \{V_1, \cdots, V_{j-1}, Z_1, \cdots, Z_j\}$. We obtain $E_{\mathscr{F}_j} I_j = p_j$ and $\Delta_1, \Delta_2, \cdots$ is a

Received October 6, 1971; revised June 16, 1972.

[1] Research partially supported by NSF Grant No. GP-23480.

martingale. Let m be a positive integer. Obviously $E_{\mathscr{F}_j}|I_j - p_j|^m \leqq 1$ and it follows, e.g. from Dharmadhikari, Fabian and Jogdeo (1968), that

$$E|\Delta_n|^m \leqq C_m n^{m/2}$$

with C_m depending only on m, not on t. The Markov inequality implies

$$(1) \qquad P\{n^{-1}|\Delta_n| \geqq n^{-r}\} \leqq C_m n^{m/2-m(1-r)} .$$

Partition R into n^r intervals with endpoints $-\infty < t_1 \leqq \cdots \leqq t_{n^r-1} < +\infty$, with mass under F_n less or equal to $2n^{-r}$ for each of these unless the interval contains one point only. It follows, using (1) in the second step, that

$$P\{n^r \sup_t |G_n(t) - F_n(t)| \geqq 3\} \leqq P\{n^r \sup_j |G_n(t_j) - F_n(t_j)| \geqq 1\}$$
$$\leqq C_m n^{m/2-m(1-r)+r} .$$

Taking $m > (1 + r)/(\frac{1}{2} - r)$ we obtain that the probability above is at most $C_m n^{-1-\eta}$ with an $\eta > 0$. It follows that with probability 1, the sequence $n^r \sup_t |G_n(t) - F_n(t)|$ is bounded. Since r was arbitrary, less than $\frac{1}{2}$, our assertion follows.

(2.2) THEOREM. *Suppose G is a distribution function with a density g, suppose g has (everywhere) a continuous derivative g' and $I(g) = \int (g'/g)^2 \, dG < +\infty$. Let $Z_1, Z_2, \cdots, V_1, V_2, \cdots$ be random variables such that $Z_n \to 0$ and each V_n is conditionally, given $Z_1, \cdots, Z_n, V_1, \cdots, V_{n-1}$, distributed according to G.*

Suppose $\Delta_n, \delta_n, \varepsilon_n$ are positive numbers such that $\Delta_n \to 0$, $\delta_n \to 0$, $\varepsilon_n \to 0$ and

$$(1) \qquad \varepsilon_n \Delta_n^{-1} \to 0 , \qquad \delta_n \varepsilon_n^{-1} \to 0 , \qquad n^{-r}\delta_n^{-1}\varepsilon_n^{-1} \to 0 \qquad \text{for an } r < \tfrac{1}{2} .$$

Let G_n be the empirical distribution function of $V_1 + Z_1, \cdots, V_n + Z_n$, i.e.

$$(2) \qquad G_n(t) = n^{-1} \sum_{j=1}^n \chi_{\{V_j+Z_j \leqq t\}} .$$

Set $d_n = G_n^{\Delta_n}$, $D_n = (G_n^{\delta_n})^{\Delta_n}$, $\kappa_n = \chi_{(\varepsilon_n, +\infty)}(d_n)$ and

$$(3) \qquad h_n(t) = \frac{D_n(t)}{2\delta_n d_n(t)} \kappa_n(t)$$

for all t in $T_n = \{(2j - 1)\Delta_n, j = 0, \pm 1, \cdots\}$ and let h_n be constant on the intervals $((2j - 2)\Delta_n, 2j\Delta_n]$. Then for P-almost all ω: The sequence $h_n(\omega, \cdot)$ converges to g'/g pointwise on $\{t; g(t) > 0\}$ and, if

$$(4) \qquad \frac{1}{n\varepsilon_n} \sum_{j=1}^n |Z_j| \to 0 ,$$

then $h_n(\omega, \cdot) \to g'/g$ in square mean with respect to dG and $\int h_n^2 \, dG_n \to I(g)$.

(2.3) EXTENSION. *Suppose the assumptions of the preceding theorem hold and η_n are functions on $\Omega \times R$, $\bar{h}_n(\omega, t) = h_n(\omega, t + \eta_n(\omega, t))$ are Borel measurable functions of t for almost all ω and $\bar{\eta}_n = \sup_t |\eta_n(\cdot, t)| \to 0$. Then for almost all ω: The sequence $\{\bar{h}_n(\omega, \cdot)\}$ converges to g'/g on $\{t; g(t) > 0\}$ and, if (2.2.4) holds and*

$$(1) \qquad \varepsilon_n^{-1}\bar{\eta}_n \to 0 , \qquad \frac{1}{n\varepsilon_n} \sum_{j=1}^n |\eta_j| \to 0$$

then $\bar{h}_n(\omega, \cdot) \to g'/g$ in square mean with respect to dG.

(2.4) EXTENSION. *Suppose the assumptions of Theorem* (2.2) *are satisfied, g is symmetric around* 0 *and g'/g is non-positive on* $[0, +\infty)$. *Let* $\bar{h}_n(t) = \frac{1}{2}(h_n(t) - h_n(-t)) \wedge 0$ *for* $t \geqq 0$, $\bar{h}_n(t) = -\bar{h}_n(-t)$ *for* $t < 0$. *Then Extension* (2.3) *remains valid if, in the definition of* $\hat{h}_n$, $\bar{h}_n$ *is used instead of* h_n.

(2.5) PROOF OF THEOREM 2.2 AND EXTENSION 2.3. It is enough to prove Extension 2.3 and the assertion $\int h_n^2\, dG_n \to I(g)$ in Theorem (2.2) since the other assertions in Theorem (2.2) follow if $\eta_n = 0$. Take an $r < \frac{1}{2}$ for which (2.2.1) is satisfied. From Lemma (2.1)

$$G_n = F_n + n^{-r}\theta_n \qquad \text{with} \quad \sup_t |\theta_n(\cdot, t)| \to 0 \,. \tag{1}$$

Let ω be a point at which all the a.e. convergences, required in the assumptions, hold. It is enough to prove all the properties of $\hat{h}_n$ at this ω. In the following all symbols denoting functions on Ω will be interpreted as the values at ω.

Let us refer to parts in (2.2.1) by a, b and c. We shall use symbols $\mathcal{O}(a_n)$, $o(a_n)$, $\mathcal{O}_u(a_n)$, $o_u(a_n)$ to denote functions h_n on R (including constants) such that $a_n^{-1}|h_n|$ are bounded, converge to zero, at each $t \in R$, uniformly on R, respectively.

From (1)

$$d_n = F_n^{\Delta_n} + o_u(n^{-r}) \tag{2}$$

and since $\varepsilon_n^{-1}n^{-r} \to 0$ by (2.2.1c),

$$\kappa_n d_n = F_n^{\Delta_n}[1 + o_u(1)]\kappa_n \,. \tag{3}$$

For every t, n let t_n be the point in T_n for which $t + \eta_n(t) \in (t_n - \Delta_n, t_n + \Delta_n]$. Suppose $g(t) > 0$. By continuity of g, if n is large enough, $f_n(\tau) \geqq \frac{1}{2}g(t)$ if $|\tau - t| < 3\Delta_n + \bar{\eta}_n$ where $f_n = F_n'$. Then $\varepsilon_n^{-1}d_n(t_n) \geqq 2\varepsilon_n^{-1}\Delta_n f_n(\tau_n) - \varepsilon_n^{-1}o_u(n^{-r}) \to +\infty$ (cf. (2.2.1a)) and $\kappa_n(t_n) = 1$, eventually.

Easy algebraic manipulation and use of (1) and (2.2.1c) and then of (3) establish

$$\hat{h}_n(t) = (1 + o_u(1))[g_n(t) + o_u(1)] \tag{4}$$

where, with χ_n an abbreviation of $\kappa_n(t_n)$,

$$g_n(t) = \frac{(F_n^{\delta_n})^{\Delta_n}(t_n)}{2\delta_n F_n^{\Delta_n}(t_n)}\, \chi_n \,. \tag{5}$$

By the mean value theorem

$$\frac{1}{2\delta_n}(F_n^{\delta_n})^{\Delta_n}(t_n) = f_n(t_n^+) - f_n(t_n^-) \tag{6}$$

with $|t_n^+ - (t_n + \Delta_n)| < \delta_n$, $|t_n^- - (t_n - \Delta_n)| < \delta_n$.

From $I(g) < +\infty$ it follows by Schwarz inequality that g' is integrable and then (cf. Rudin (1966) Theorem 8.21) g is bounded (also absolutely continuous) and so are, uniformly, all f_n. It is then easy to obtain, using (3) and (2.2.1b), that

$$\chi_n F_n^{\Delta_n}(t_n) = [F_n(t_n^+) - F_n(t_n^-)][1 + o_u(1)]\chi_n \,. \tag{7}$$

From (5), (6), (7)

$$g_n^2(t) = \chi_n \frac{[f_n(t_n^+) - f_n(t_n^-)]^2}{F_n(t_n^+) - F_n(t_n^-)} [F_n^{\Delta_n}(t_n)]^{-1}(1 + o_u(1)) \tag{8}$$

and, as $t_n^+ \to t$, $t_n^- \to t$, $Z_i \to 0$ and g', g are continuous, we obtain

$$g_n(t) \to (g'/g)(t)\,, \qquad \tilde{h}_n(t) \to (g'/g)(t) \qquad \text{if} \quad g(t) > 0\,. \tag{9}$$

Now we shall prove the $L_2(g)$ convergence under the additional condition (2.2.4) and (2.3.1). In view of (9) and (4) it is enough to prove the uniform integrability of g_n^2 and that will follow if we prove $\limsup \int g_n^2\, dG \leqq \int (g'/g)^2\, dG$ (use the Fatou lemma, and Theorem C, page 163 in Loève (1963)).

Starting with the numerator in (8) we use a simple identity and the Schwarz inequality to obtain

$$[f_n(t_n^+) - f_n(t_n^-)]^2 = [\textstyle\int_{t_n^-}^{t_n^+} (f_n'/f_n)\, dF_n]^2 \leqq (F_n(t_n^+) - F_n(t_n^-)) \int_{t_n^-}^{t_n^+} (f_n'/f_n)^2\, dF_n\,.$$

Also, since g is bounded by a constant K,

$$F_n^{\Delta_n} = G^{\Delta_n} + R_n$$

where $\varepsilon_n^{-1}|R_n| \leqq 2K/(n\varepsilon_n) \sum_{j=1}^n |Z_j| \to 0$ by (2.2.4).

Hence, using (3) and the definition of κ_n, we conclude that

$$\kappa_n F_n^{\Delta_n} = \kappa_n G^{\Delta_n}[1 + o_u(1)] \tag{10}$$

and, from (8),

$$g_n^2(t) \leqq \chi_n[1 + o_u(1)][G^{\Delta_n}(t_n)]^{-1} \int_{t_n^-}^{t_n^+} (f_n'/f_n)^2\, dF_n\,. \tag{11}$$

For n sufficiently large, $\bar{\eta}_n < \Delta_n$ by (2.3.1) and (2.2.1a) and then if t is in $((2j-2)\Delta_n, 2j\Delta_n)$, t_n is $(2j-1)\Delta_n$ if the distance of t to the complement of the interval is greater than $\bar{\eta}_n$; otherwise t_n can also be either $(2j-3)\Delta_n$ or $(2j+1)\Delta_n$. Thus, denoting $\tau_{n,j} = (2j-1)\Delta_n$,

$$\int_{\tau_{n,j}-\Delta_n}^{\tau_{n,j}+\Delta_n} g_n^2\, dG \leqq [1 + o_u(1)] \Big\{ \int_{\tau_{n,j}^-}^{\tau_{n,j}^+} (f_n'/f_n)^2\, dF_n + \frac{\mathcal{O}_u(\bar{\eta}_n)}{\varepsilon_n} \Big[\int_{\tau_{n,j-1}^-}^{\tau_{n,j-1}^+} (f_n'/f)^2\, dF_n + \int_{\tau_{n,j+1}^-}^{\tau_{n,j+1}^+} (f_n'/f_n)^2\, dF_n \Big] \Big\}$$

and subsequently, by (2.3.1),

$$\int g_n^2\, dG \leqq [1 + o_u(1)] \sum_{j=-\infty}^{+\infty} \int_{\tau_{n,j}^-}^{\tau_{n,j}^+} (f_n'/f_n)^2\, dF_n$$

and

$$\int g_n^2\, dG \leqq [1 + o_u(1)][I(f_n) + 2 \textstyle\int_{B_n} (f_n'/f_n)^2\, dF_n] \tag{12}$$

with

$$B_n = \bigcup \{(2j\Delta_n - \delta_n, 2j\Delta_n + \delta_n); j = 0, \pm 1, \cdots\}\,.$$

Using the Schwarz inequality and the fact that $g(t) = 0$ implies $g'(t) = 0$;

$$(f_n'(t))^2 = \Big(\sum_{i=1}^n \frac{g'(t - Z_i)}{(g(t - Z_i))^{\frac{1}{2}}} (g(t - Z_i))^{\frac{1}{2}} \frac{1}{n} \Big)^2 \leqq \frac{1}{n} \sum_{i=1}^n \frac{g'^2(t - Z_i)}{g(t - Z_i)} f_n(t)$$

we conclude, first, that $I(f_n) \leqq I(g)$, and, secondly

$$\int_{B_n} (f_n'/f_n)^2 \, dF_n \leqq \frac{1}{n} \sum_{i=1}^n \int_{B_n - Z_i} \left(\frac{g'}{g}\right)^2 dG$$
$$\leqq \frac{1}{n} \sum_{i=1}^n \int_{(B_n - Z_i) \cap (-K, K)} \left(\frac{g'}{g}\right)^2 dG + \varepsilon$$

if ε is positive, K suitably chosen and $B_n - Z_i$ denotes B_n shifted by $-Z_i$. The Lebesgue measure of each of these sets is at most $2\delta_n(K\Delta_n^{-1} + 3) = o(1)$ and thus the integrals converge to zero. This implies

$$\limsup \int g_n^2 \, dG \leqq I(g)$$

and completes the proof of the $L_2(g)$ convergence.

It remains to prove $\int h_n^2 \, dG_n \to I(g)$. Because of (4) (with $\eta_n = 0$) it is enough to prove $\int g_n^2 \, dG_n \to I(g)$. We know that $\int g_n^2 \, dG \to I(g)$. Write

$$\int g_n^2 \, dG = \sum_{j=-\infty}^{+\infty} \int_{\tau_{n,j}-\Delta_n}^{\tau_{n,j}+\Delta_n} g_n^2(\tau_{n,j}) G^{\Delta_n}(\tau_{n,j}) \, .$$

Notice that $\chi_n G^{\Delta_n} = \chi_n[1 + o_u(1)]G_n^{\Delta_n}$ by (3) and (10) and thus $\int g_n^2 \, dG_n = [1 + o_u(1)] \int g_n^2 \, dG$. This completes the proof.

(2.6) PROOF OF EXTENSION (2.4). If $t \geqq 0$ and $g(t) > 0$ then $h_n(t + \eta_n(t)) \to (g'/g)(t) \geqq 0$, $-h_n(-t - \eta_n) \to -(g'/g)(-t) = g'/g(t)$ and $\bar{h}_n(t + \eta_n(t)) \to (g'/g)(t)$. Thus the new $\bar{h}_n$ converge to g'/g on $\{t;\, g(t) > 0\}$ again. Obviously $\bar{h}_n^2(t + \eta_n(t)) \leqq h_n^2(t + \eta_n(t)) + h_n^2(-t - \eta_n(t))$ and the uniform integrabilily with respect to dG follows since g is symmetric and $h_n^2(t + \eta_n(t))$ are uniformly integrable by Extension (2.3).

REFERENCES

DHARMADHIKARI, S. W., FABIAN, V. and JOGDEO, K. (1968). Bounds on the moments of martingales. *Ann. Math. Statist.* **39** 1719–1723.

FABIAN, V. (1973). Asymptotically efficient stochastic approximation; the RM case. *Ann. Statist.* **1** 486–495.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Statistics.* Academia, Prague.

LOÈVE, M. (1963). *Probability Theory*, (3rd ed.). Van Nostrand, Princeton.

RUDIN, W. (1966). *Real and Complex Analysis.* McGraw Hill, New York.

SCHUSTER, E. F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* **40** 1187–1195.

VAN EEDEN, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.* **41** 172–181.

WEISS, L. and WOLFOWITZ, J. (1970). Asymptotically efficient non-parametric estimators of location and scale parameters. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **16** 134–150.

DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
WELLS HALL
EAST LANSING, MICHIGAN 48823