Z. Wahrscheinlichkeitstheorie verw. Geb. 14, 323–330 (1970)

A Characterization of Limiting Distributions of Regular Estimates

JAROSLAV HÁJEK

Summary. We consider a sequence of estimates in a sequence of general estimation problems with a k-dimensional parameter. Under certain very general conditions we prove that the limiting distribution of the estimates, if properly normed, is a convolution of a certain normal distribution, which depends only of the underlying distributions, and of a further distribution, which depends on the choice of the estimate. As corollaries we obtain inequalities for asymptotic variances and for asymptotic probabilities of certain sets, generalizing so some results of J. Wolfowitz (1965), S. Kaufman (1966), L. Schmetterer (1966) and G. G. Roussas (1968).

1. Introduction

In several papers there were established asymptotic lower bounds for variances and for probabilities of some sets provided the estimates are regular enough to avoid superefficiency. See Rao (1963), Bahadur (1964), Wolfowitz (1965), Kaufman (1966), Schmetterer (1966), Roussas (1968), for example. In the present paper we obtain some of their results as corollaries of a representation theorem describing the class of all possible limiting laws. The present result refers to a general sequence of experiments with general norming matrices K_n. Our condition (2) below is implied by conditions imposed on the families of distributions in the above mentioned papers. The same is true about regularity condition (3) concerning the limiting distribution of estimates. A comparison with Kaufman (1966) is given in Remark 1 below, and the same comparison could be made with Wolfowitz (1965). Schmetterer's (1966) condition, namely that the distribution law of normed estimates converges continuously in θ, also entails (3). See also Remark 2 for Bahadur's (1964) condition, which is weaker than (3).

Roughly speaking, condition (2) means that the likelihood functions may be locally asymptotically approximated by a family of normal densities differing in location only. Similarly, condition (3) expresses that the family of distributions of an estimate T_n may be approximated, in a local asymptotic sense, by a family of distributions differing again in location only.

The idea of the proof is based on considering the parameter as a random vector that has uniform distribution over certain cubes. In this respect the spirit of the proof is Bayesian, similarly as in Weiss and Wolfowitz (1967). The mathematical technique of the present paper is borrowed from LeCam (1960).

2. The Theorem

Consider a sequence of statistical problems $(\mathscr{X}_n, \mathscr{A}_n, P_n(\cdot, \theta))$, $n \geq 1$, where θ runs through an open subset Θ of R^k. The nature of sets $\mathscr{X}_n$ may be completely arbitrary, and the only connection between individual statistical problems will be given by conditions (2) and (3). In particular, the n-th problem may deal with n observations which may be independent, or may form a Markov chain, or whatsoever. A point from $\mathscr{X}_n$ will be denoted by x_n, and X_n will denote the abstract random variable defined by the identity function $X_n(x_n)=x_n$, $x_n\in\mathscr{X}_n$. The norm of a point $h=(h_1,\dots,h_k)$ from R^k will be defined by $|h|=\max_{1\leq i\leq k}|h_i|$; $h\leq v$ will mean that $h_i\leq v_i$, $1\leq i\leq k$, and $h'v$ will denote the scalar product of two vectors $h, v\in R^k$.

Take a point $t\in\Theta$ and assume it is the true value of the parameter θ. We shall abbreviate $P_n=P_n(\cdot,t)$ and $P_{nh}=P_n(\cdot,t+K_n^{-1}h)$ where K_n^{-1} are inverses of some norming regular $(k\times k)$-matrices. Most usually the theorem below is applicable with K_n diagonal and having $n^{\frac{1}{2}}$ for its diagonal elements; then simply $K_n^{-1}h=n^{-\frac{1}{2}}h$. We shall need that $\{K_n^{-1}\}$ be a sequence of contracting transforms in the sense that for every $\varepsilon>0$ and $a>0$ there exists an n_0 such that $|h|<a$ entails $|K_n^{-1}h|<\varepsilon$ for all $n>n_0$. This property will be insured by

$$\lim_{n\to\infty}(\text{the minimal eigenvalue of } K_n' K_n)=\infty, \tag{1}$$

where K_n' denotes the transpose of K_n. Since Θ is open, (1) ensures that for all $h\in R^k$ the points $t+K_n^{-1}h$ belong to Θ if n is sufficiently large. If $\Theta=R^k$, the theorem below makes sense and is true without condition (1).

Given two probability measures P and Q, denote by dQ/dP the Radon-Nikodym derivative of the *absolutely continuous part* of Q with respect to P. Thus, generally, we do not obtain $\int_A (dQ/dP)\,dP=Q(A)$, but only $\leq Q(A)$. Introduce the familiy of likelihood ratios

$$r_n(h,x_n)=\frac{dP_{nh}}{dP_n}(x_n), \qquad h\in R^k,\ n\geq n_h,\ x_n\in\mathscr{X}_n,$$

where n_h denotes the smallest n_0 such that $n\geq n_0$ entails $t+K_n^{-1}h\in\Theta$. In what follows the argument x_n will be usually omitted in order to stress the other arguments.

The distribution law of a random vector $Y_n=Y_n(x_n)$ under P_n will be denoted by $\mathscr{L}(Y_n|P_n)$ and its weak convergence to L by $\mathscr{L}(Y_n|P_n)\to L$. The k-dimensional normal distribution with zero expectation vector and covariance matrix Γ will be denoted by $\Phi(\cdot|\Gamma)$, and the corresponding law by $\mathscr{N}(0,\Gamma)$. The expectation with respect P_n will be denoted by E_n, i.e. $E_n(\cdot)=\int(\cdot)\,dP_n$.

Let $T_n=T_n(x_n)$ be a sequence of estimates of θ, i.e. for every n T_n is a measurable transform from $\mathscr{X}_n$ to R^k. (We shall say that the estimates T_n are regular, if they satisfy condition (3) below.) Now we are prepared to formulate the following

Theorem. *In the above notations, let us assume that* (1) *holds and that*

$$r_n(h)=\exp\{h'\Delta_n-\tfrac{1}{2}h'\Gamma h+Z_n(h)\}, \qquad h\in R^k,\ n\geq n_h, \tag{2}$$

where $\mathscr{L}(\Delta_n|P_n)\to\mathscr{N}(0,\Gamma)$ and $Z_n(h)\to 0$ in P_n-probability for every $h\in R^k$. Further consider a sequence of estimates satisfying

$$P_{nh}(K_n(T_n-t)-h\leq v)\to L(v), \qquad \textit{for every } h\in R^k, \tag{3}$$

in continuity points of some distribution function $L(v)$, $v\in R^k$.

Then, if the matrix Γ is regular, we have

$$L(v)=\int \Phi(v-u|\Gamma^{-1})\,dG(u), \tag{4}$$

where $G(u)$ is a certain distribution function in R^k.

Remark. The dependence of Δ_n, Γ and $Z_n(h)$ on t, and of Δ_n and $Z_n(h)$ on x_n, is suppressed in our notation.

Proof. By Lemma 1 below we may also write

$$r_n(h)=\exp\{h'\Delta_n^*-\tfrac{1}{2}h'\Gamma h+Z_n^*(h)\} \tag{5}$$

where the properties $\mathscr{L}(\Delta_n^*|P_n)\to\mathscr{N}(0,\Gamma)$ and $Z_n^*(h)\to 0$ in P_n-probability are preserved, and, in addition

$$B_{nh}=[E_n\exp(h'\Delta_n^*-\tfrac{1}{2}h'\Gamma h)]^{-1} \tag{6}$$

satisfies

$$\sup_{|h|<j}|B_{nh}-1|\to 0 \qquad \text{for every } j>0, \text{ as } n\to\infty. \tag{7}$$

Put

$$r_n^*(h)=B_{nh}\exp(h'\Delta_n^*-\tfrac{1}{2}h'\Gamma h). \tag{8}$$

From our assumptions and from (7) it follows that $N(-\frac{1}{2}h'\Gamma h, h'\Gamma h)$ is a limit for $\mathscr{L}(\log r_n(h)|P_n)$ as well as for $\mathscr{L}(\log r_n^*(h)|P_n)$. Let Y be a random variable such that $\mathscr{L}(Y)=\mathscr{N}(-\frac{1}{2}h'\Gamma h, h'\Gamma h)$. Then $E\,e^Y=1$. Further, by (8) and (6),

$$E_n\,r_n^*(h)=1.$$

On the other hand, since $r_n(h)$ is a the Radon-Nikodym derivative of the absolutely continuous part of P_{nh} relative to P_n,

$$\limsup_n E_n\,r_n(h)\leqq 1.$$

Since $r_n(h)$ is nonnegative and convergent in distribution to e^Y, we also have

$$\liminf_n E\,r_n(h)\geqq e^Y=1.$$

Consequently,

$$\lim_{n\to\infty}E_n\,r_n(h)=1.$$

Since, furthermore, $[r_n(h)-r_n^*(h)]\to 0$ in P_n-probability, Lemma 2 below may be applied to the effect that

$$\int|r_n(h)-r_n^*(h)|\,dP_n\to 0, \qquad h\in R^k. \tag{9}$$

Put

$$Q_{nh}(A)=\int_A r_n^*(h)\,dP_n, \qquad A\in\mathscr{A}_n. \tag{10}$$

Then (3) and (9) entail that

$$Q_{nh}(K_n(T_n-t)-h\leqq v)\to L(v). \tag{11}$$

From (8) it follows that $Q_{nh}(K_n(T_n-t)-h\leqq v)$ is a measurable function of h for every $v\in R^k$.

Let $\lambda_j(dh)$ be the uniform distribution over the cube $|h| \leqq j$. It is easy to see that (11) entails for every natural j

$$\bar{L}_{nj}(v) \overset{\text{Df}}{=} \int Q_{nh}(K_n(T_n - t) - h \leqq v)\, d\lambda_j(h) \to L(v). \tag{12}$$

By a LeCam's lemma (LeCam, 1960 or Hájek-Šidák, 1967, VI.1.4) we deduce

$$\mathscr{L}(\Gamma^{-1} \Delta_n^* | Q_{nh}) \to \mathscr{N}(h, \Gamma^{-1}). \tag{13}$$

Put $a_j = (-j, \ldots, -j)$, $b_j = (j, \ldots, j)$ and $c_j = (\sqrt{j}, \ldots, \sqrt{j})$. Let U be some random k-vector possessing normal distribution with zero expectation and covariance matrix Γ^{-1}, i.e. $\mathscr{L}(U) = \mathscr{N}(0, \Gamma^{-1})$. Then, by (13),

$$\begin{aligned} &\int Q_{nh}(a_j + c_j \leqq \Gamma^{-1} \Delta_n^* \leqq b_j - c_j)\, d\lambda_j(h) \\ &\quad \to \int P(a_j + c_j \leqq U + h \leqq b_j - c_j)\, d\lambda_j(h) \\ &\quad \geqq \int P(a_j + 2c_j \leqq h \leqq b_j - 2c_j)\, d\lambda_j(h) - P(|U| > \sqrt{j}) \\ &\quad = (1 - 2/\sqrt{j})^k - P(|U| > \sqrt{j}). \end{aligned} \tag{14}$$

For every natural j denote by m_j some integer such that

$$\sup_{|h| < j} |B_{nh} - 1| < \frac{1}{j}, \qquad n > m_j, \tag{15}$$

$$\rho(\bar{L}_{nj}, L) < \frac{1}{j}, \qquad n > m_j \tag{16}$$

(where $\bar{L}_{nj}$ is given by (12) and ρ denotes the Lévy distance), and such that

$$\begin{aligned} &\int Q_{nh}(a_j + c_j \leqq \Gamma^{-1} \Delta_n^* \leqq b_j - c_j)\, d\lambda_j(h) \\ &\quad > (1 - 2/\sqrt{j})^k - P(|U| > \sqrt{j}) - 1/j, \qquad n > m_j. \end{aligned} \tag{17}$$

All this may be satisfied in view of (7), (12) and (14). We may assume that $m_1 < m_2 < \cdots$. Let $j(n)$ be defined by

$$m_{j(n)} \leqq n < m_{j(n)+1},$$

and set

$$\bar{L}_n = \bar{L}_{n\,j(n)}, \qquad \bar{Q}_n = \int Q_{nh}\, d\lambda_{j(n)}(h). \tag{18}$$

From (16) it follows that

$$\bar{L}_n \to L. \tag{19}$$

On the other hand $\bar{L}_n(v)$ may also be interpreted as the probability of $\{K_n(T_n - t) - h \leqq v\}$, if the joint distribution of (X_n, h) is given by the prior distribution $\lambda_{j(n)}$ and by the family Q_{nh} of conditional distributions of X_n given h. Denoting the posterior distribution function of h by $D_n(y | X_n)$, we may therefore write

$$\bar{L}_n(v) = \int [1 - D_n(K_n(T_n - t) - v - 0 | X_n)]\, d\bar{Q}_n. \tag{20}$$

Now, in view of (8), for $a_{j(n)} \leqq y \leqq b_{j(n)}$,

$$\begin{aligned} D_n(y|X_n) &= c(X_n) \int_{a_{j(n)}}^{y} B_{nh} \exp(h' \Delta_n^* - \tfrac{1}{2} h' \Gamma h)\, dh \\ &= c'(X_n) \int_{a_{j(n)}}^{y} B_{nh} \exp\{(h - \Gamma^{-1} \Delta_n^*)' \Gamma (h - \Gamma^{-1} \Delta_n^*)\}\, dh \end{aligned} \tag{21}$$

where $c(X_n)$ and $c'(X_n)$ are some constants depending on X_n only. Now, in view of (15),

$$\sup_{|h| < j(n)} |B_{nh} - 1| \to 1, \tag{22}$$

and in view of (17),

$$-\infty \leftarrow a_{j(n)} - \Gamma^{-1} \Delta_n^* \leqq h - \Gamma^{-1} \Delta_n^* \leqq b_{j(n)} - \Gamma^{-1} \Delta_n^* \to \infty \tag{23}$$

where $-\infty = (-\infty, \ldots, -\infty)$, $\infty = (\infty, \ldots, \infty)$ and the convergence is in $\bar{Q}_n$-probability. Consequently

$$D_n(y|X_n) \to \Phi(y - \Gamma^{-1} \Delta_n^* | \Gamma^{-1}) \tag{24}$$

in $\bar{Q}_n$-probability, and, in turn

$$\left|\bar{L}_n(v) - \int \Phi\big(v - K_n(T_n - t) + \Gamma^{-1} \Delta_n^* | \Gamma^{-1}\big)\, d\bar{Q}_n\right| \to 0. \tag{25}$$

Denoting

$$G_n(u) = \bar{Q}_n\big(K_n(T_n - t) - \Gamma^{-1} \Delta_n^* \leqq u\big) \tag{26}$$

we may also write

$$\left|\bar{L}_n(v) - \int \Phi(v - u | \Gamma^{-1})\, dG_n(u)\right| \to 0. \tag{27}$$

Taking a subsequence $\{m\} \subset \{n\}$ such that $G_m \to G$, we obtain from (19) and (27)

$$L(v) \leftarrow \bar{L}_m(v) \to \int \Phi(v - u | \Gamma^{-1})\, dG(u).$$

Since $L(v)$ is a distribution function, $G(u)$ has to be a distribution function, too. This completes the proof.

Lemma 1. *Let $\mathcal{L}(\Delta_n | P_n) \to \mathcal{N}(0, \Gamma)$ hold. Then there exists a truncated version Δ_n^* of Δ_n such that $(\Delta_n - \Delta_n^*) \to 0$ in P_n-probability and*

$$\sup_{|h| < j} |E_n \exp(h' \Delta_n^* - \tfrac{1}{2} h' \Gamma h) - 1| \to 0, \qquad j > 0. \tag{28}$$

Proof. For every natural i let us put

$$\begin{aligned} \Delta_{ni} &= \Delta_n, \qquad &&\text{if } |\Delta_n| \leqq i, \\ &= 0, &&\text{otherwise.} \end{aligned}$$

Let Y be some random vector such that $\mathcal{L}(Y) = \mathcal{N}(0, \Gamma)$ and let $Y_i = Y$, if $|Y| \leqq i$, and 0 otherwise. Let m_i be an integer such that

$$\sup_{|h| \leqq i} |E_n \exp(h' \Delta_{ni} - \tfrac{1}{2} h' \Gamma h) - E \exp(h' Y_i - \tfrac{1}{2} h' \Gamma h)| < \frac{1}{i}, \qquad n > m_i.$$

Such an m_i exists, since $\mathscr{L}(\Delta_{ni}|P_n) \to \mathscr{L}(Y_i)$ and the distributions are concentrated on a compact subset on which the system of functions $\exp(h'y - \frac{1}{2} h \Gamma h)$ is compact in the supremum metric. Note that further

$$\lim_{i\to\infty} \sup_{|h|<j} |E \exp(h' Y_i - \tfrac{1}{2} h' \Gamma h) - 1| = 0.$$

Assume that $m_1 < m_2 < \cdots$, define $i(n)$ by $m_{i(n)} \leqq n < m_{i(n)+1}$, and put $\Delta_n^* = \Delta_{ni(n)}$. Then the conclusion easily follows.

Lemma 2. *Let $\{(U_n, V_n), n \geqq 1\}$ be a sequence of pairs of nonnegative random variables such that*

(a) *U_n converges in distribution to U, $EU < \infty$.*

(b) *$(U_n - V_n) \to 0$ in probability.*

(c) *$E_n U_n \to EU$, $E_n V_n \to EU$, where E_n and E denote expectations.*

Then

$$E_n |U_n - V_n| \to 0 \qquad \text{as } n \to \infty. \tag{29}$$

Proof. The proof follows from Loève (1963), Theorem 11.4.A.

3. Corollaries

Corollary 1. *Let C be a convex symmetric set in R^k, and let Y be a random variable such that $\mathscr{L}(Y) = \mathscr{N}(0, \Gamma^{-1})$. Then, if the assumptions of the above theorem are satisfied for some $t \in \Theta$, we have for T_n from the same theorem*

$$\limsup_{n\to\infty} P_n[K_n(T_n - t) \in C | t] \leqq P(Y \in C). \tag{30}$$

Proof. Since the boundary of C has zero Lebesgue measure and since the limiting law of $K_n(T_n - t)$ is absolutely continuous, in view of (4), we have

$$\lim_{n\to\infty} P_n[K_n(T_n - t) \in C | t] = P(Z \in C),$$

where $\mathscr{L}(Z) = L$. Furthermore, (4) with the well-known lemma of Anderson (1955) entail

$$P(Z \in C) \leqq P(Y \in C).$$

Remark 1. Corollary 1 provides the essence of the main result by Kaufman (1966). His conditions, namely that the distribution of $K_n(T_n - \theta)$ converges uniformly in $R^k \times C$ where C is any compact subset of Θ, and his Lemma 5.1 entail our assumption (3). His regularity conditions for densities entail assumption (2). Contrary to Kaufman, we do not claim that the maximum likelihood estimate is after norming asymptotically normal $\mathscr{N}(0, \Gamma^{-1})$. For this conclusion we would need some global conditions concerning the distributions.

Under LeCam's *DN* conditions in [4], there is an estimate, not necessarily the maximum likelihood one, which is after norming asymptotically normal $\mathscr{N}(0, \Gamma^{-1})$.

Remark 2. The theorem also entails

$$\liminf_{n\to\infty} E_n(h' K_n(T_n - t))^2 \geqq h' \Gamma^{-1} h. \tag{31}$$

If we assume that $K_n(T_n-\theta)$ is asymptotically normal with zero variance, then by Bahadur (1964), we could derive (31) from weaker asumptions than (3). Actually, then (3) may be replaced by

$$P_{nh}\big(h' K_n(T_n-t)\leqq h' h\big)\to\tfrac{1}{2}, \qquad h\in R^k. \tag{32}$$

Remark 3. We have introduced K_n instead of $n^{\frac{1}{2}}I$ in order to cover such instances as

$$p_n(\theta)=\prod_{i=1}^{n} f\left(x_i-\sum_{j=1}^{k}\theta_j c_{nij}\right) \tag{33}$$

where we may put $K_n=\left\{\sum_{i=1}^{n} c_{nij}\, c_{nij'}\right\}_{j,j'=1}^{k}$. Alternatively, we may take for K_n a diagonal matrix coinciding on the diagonal with the previous one. Instances of this character occur in asymptotic theory of linear regression.

Corollary 2. *Put $A_v=\{y\colon a' y\leqq v\}$ where $a\in R^k$ and $v\in R$. Assume that*

$$\int_{A_v} dL(y)=\int_{A_{v^*}} d\Phi(y|\Gamma^{-1}). \tag{34}$$

Then, under the assumptions of the above theorem

$$\int_{A_{v+s}\div A_v} dL(y)\leqq \int_{A_{v^*+s}\div A_{v^*}} d\Phi(y|\Gamma^{-1}), \qquad s\in R, \tag{35}$$

where $\div$ denotes the symmetric difference.

Proof. Consider two distributions of the pair $(Y, U)\in R^k\times R^k$, namely

$$\Phi(\cdot|\Gamma^{-1})\times G \quad \text{and} \quad \Phi(\cdot+\lambda\Gamma^{-1} a|\Gamma^{-1})\times G, \qquad \lambda>0.$$

Then (34) entails that the tests rejecting $\Phi(\cdot|\Gamma^{-1})\times G$ if $a' Y\leqq v^*$ and $a'(Y+U)\leqq v$, respectively, have the same significance level, and (35) simply means that the former test has for $\lambda a'\Gamma^{-1} a=s>0$ a larger power than the latter test, which easily follows from the Neyman-Pearson lemma. For $s<0$ we proceed similarly.

Remark 4. Corollary 2 generalizes a result by Wolfowitz (1965) and Roussas (1968), if we note that

$$P_n\big(a' K_n(T_n-t)\leqq v\big)\to\int_{A_v} dL(v),$$

which follows again from the fact the A_v has zero boundary.

Remark 5. If the loss incurred by T_n if θ is true equals $l\big(K_n(T_n-\theta)\big)$, where $l(\cdot)$ may be represented as a mixture of indicators of complements of convex symmetric sets, then Corollary 1 entails that

$$\liminf_{n\to\infty} E_n\, l_n[K_n(T_n-t)]\geqq E\, l(Y),$$

where $\mathscr{L}(Y)=\mathscr{N}(0,\Gamma^{-1})$. In this sense $\mathscr{N}(0,\Gamma^{-1})$ may be regarded as a best possible limiting distribution. From (26) it is apparent that $\mathscr{L}\big(K_n(T_n-t)|P_n\big)\to\mathscr{N}(0,\Gamma^{-1})$ if and only if $[K_n(T_n-t)-\Gamma^{-1}\Delta_n^*]\to 0$ in $\bar{Q}_n$-probability.

References

1. Anderson, T. W.: The integral of a symmetric unimodal function. Proc. Amer. math. Soc. **6**, 170 – 176 (1955).
2. Bahadur, R. R.: A note on Fisher's bound for asymptotic variance. Ann. math. Statistics **35**, 1545 – 1552 (1964).
3. Hájek, J., Šidák, Z.: Theory of rank tests. Academia, Prague, London: Academic Press 1967.
4. LeCam, L.: Locally asymptotically normal families of distributions. Univ. California Publ. Statist., 1960.
5. Kaufman, S.: Asymptotic efficiency of the maximum likelihood estimator. Ann. Inst. statist. Math. **18**, 155 – 178 (1966).
6. Loève, M.: Probability theory, 3rd ed. Princeton-New Jersey-Toronto-New York-London: D. Van Nostrand 1963.
7. Rao, C. R.: Criteria of estimation in large samples. Sankhya **25** a, 189 – 206 (1963).
8. Roussas, G. G.: Some applications of the asymptotic distribution of likelihood functions to the asymptotic efficiency of estimates. Z. Wahrscheinlichkeitstheorie verw. Geb. **10**, 252 – 260 (1968).
9. Schmetterer, L.: On asymptotic efficiency of estimates. Research papers in statistics, F. N. David (editor). New York: Wiley 1966.
10. Weiss, L., Wolfowitz, J.: Generalized maximum likelihood estimators. Teor. Verojatn. Primen **11**, 68 – 93 (1966).
11. – Maximum probability estimators. Ann. Inst. statist. Math. **19**, 103 – 206 (1967).
12. Wolfowitz, J.: Asymptotic efficiency of the maximum likelihood estimator. Teor. Verojatn. Primen **10**, 267 – 281 (1965).

Professor J. Hájek
Department of Statistics
The Florida State University
Tallahassee 32306, Florida, USA

(Received November 21, 1968)