In the cases presented above, the dimension, m, of the parameter space ω of the null hypothesis is either 0 or 1. This can be extended somewhat. If the null hypothesis is that the θ's fall into m equal sets, $-2 \ln \lambda$ is distributed as χ^2 with $2(k - m)$ degrees of freedom provided the null hypothesis is true. For example, suppose $k = 6$ and that we test the hypothesis $\theta_1 = \theta_2 = \theta_3 = \theta_4$ and $\theta_5 = \theta_6$ against all possible alternatives. Then $-2 \ln \lambda$ has a χ^2-distribution with $2(6 - 2) = 8$ degrees of freedom.

REFERENCES

[1] R. C. Davis, "On minimum variance in nonregular estimation," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 43–57.
[2] J. Neyman, "On statistics the distribution of which is independent of the parameters involved in the original probability law of the observed variables," *Statist. Res. Mem. London*, Vol. 2 (1938), pp. 58–59.
[3] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypothesis," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 60–62.

AN APPLICATION OF CHUNG'S LEMMA TO THE KIEFER-WOLFOWITZ STOCHASTIC APPROXIMATION PROCEDURE[1]

By Cyrus Derman[2]

Syracuse University

1. Summary. Let $M(x)$ be a strictly increasing regression function for $x < \theta$, and strictly decreasing regression function for $x > \theta$. Under conditions 1, 2, and 3 given below, the stochastic approximation procedure proposed by Kiefer and Wolfowitz [3] is shown to converge stochastically to θ. Under the additional conditions 4, 5, 6 given below, the procedure is shown to converge in distribution to the normal distribution. Our method is the one used by Chung [2].

2. Introduction. Let $H(y \mid x)$ be a family of distribution functions which depend on the parameter x and let $M(x) = \int_{-\infty}^{\infty} y \, dH(y \mid x)$. Suppose $M(x)$ is strictly increasing for $x < \theta$, and strictly decreasing for $x > \theta$. Let $\{a_n\}$ and $\{c_n\}$ be sequences of positive numbers such that

$$c_n \to 0, \qquad \sum a_n = \infty, \qquad \sum a_n c_n < \infty, \qquad \sum a_n^2 c_n^{-2} < \infty.$$

Kiefer and Wolfowitz [3] suggested a recursive scheme for estimating θ which is as follows. Let z_1 be an arbitrary real number. For all positive integral n,

$$Z_{n+1} = Z_n + \frac{a_n}{c_n} (y_{2n} - y_{2n-1}), \tag{1}$$

where y_{2n-1} and y_{2n} are independent chance variables with respective distributions $H(y \mid z_n + c_n)$ and $H(y \mid z_n - c_n)$. Under certain regularity conditions

Received March 21, 1955.
[1] This research was supported in part by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.
[2] Now at Columbia University.

they proved that Z_n converges stochastically to θ as $n \to \infty$. Blum [1], later, under slightly weaker conditions proved convergence with probability one. However, the regularity conditions imposed are such, that they are not satisfied if $M(x) = K - K'(x - \theta)^2$ where K and K' are any constants ($K' > 0$). Since this case is a common one, it seems important that it be considered. Below, we shall prove some convergence theorems under conditions which are quite restrictive. However, the above function, and functions of a slightly more general type, will satisfy these conditions. Since our purpose is to show that the Kiefer-Wolfowitz procedure is applicable in the parabolic case, no attempt will be made here to weaken the conditions.

The main tool to be used is the following lemma proved by Chung [2] which he used in his analysis of the Robbins-Monrò [4] procedure.

LEMMA. *Suppose* $\{b_n\}, n \geqq 1$ *is a sequence of real numbers such that for* $n \geqq n_0$,

$$b_{n+1} \geqq (1 - c / n^s)\, b_n + c_1 / n^t, \quad \text{where } 0 < s < 1,\ s < t,\ c > 0,\ c_1 > 0.$$

Then $\underline{\lim}_{n\to\infty} n^{t-s} b_n \geqq c_1 / c$. If

$$b_{n+1} \leqq \left(1 - \frac{c_n}{n^s}\right) b_n + \frac{c_2}{n^t} \tag{2}$$

where $0 < s < 1$, $s < t$, $c_n \geqq c > 0$, $c_2 < 0$, *then* $\overline{\lim}_{n\to\infty} n^{t-s} b_n \leqq c_2 / c$.

We remark that if in (2) c_2 is replaced by a sequence $\{c_{2n}\}$ of positive numbers such that $c_{2n} \to 0$, then $\overline{\lim}_{n\to\infty} n^{t-s} b_n \leqq 0$.

3. A Convergence Theorem. We postulate the following conditions.

CONDITION 1. *There exist positive constants* K_1, K_2, *and* C_0 *such that for every* c, *where* $0 < c < C_0$

$$-cK_2(x - \theta)^2 \leqq (M(x + c) - M(x - c))(x - \theta) \leqq -cK_1(x - \theta)^2.$$

The above is a condition that the function $M(x)$ does not increase (decrease) too rapidly or too slowly. We remark that for $M(x) = K - K'(x - \theta)^2$, $K_1 = K_2 = K'$.

CONDITION 2. *Let* $\sigma^2(x) = \int_{-\infty}^{\infty} (y - M(x))^2\, dH(y|x)$. *There exist real numbers* M_1 *and* M_2 *such that*

$$0 < M_1 \leqq \sigma^2(x) \leqq M_2 < \infty.$$

CONDITION 3. *Let* η *and* ϵ *be any two real numbers such that* $\eta > \epsilon > 0$ *and* $\eta + \epsilon < \frac{1}{2}$. *We set* $a_n = 1 / n^{1-\epsilon}$ *and* $c_n = 1 / n^{1/2-\eta}$.

THEOREM 1. *If Conditions* 1, 2 *and* 3 *hold, then*

$$\frac{M_1}{K_2} \leqq \underline{\lim}_{n\to\infty} n^{2\eta-\epsilon} b_n \leqq \overline{\lim}_{n\to\infty} n^{2\eta-\epsilon} b_n \leqq \frac{M_2}{K_1}, \qquad \text{where } b_n = E(Z_n - \theta)^2. \tag{3}$$

PROOF. From (1) we have

$$b_{n+1} = b_n + \frac{2a_n}{c_n} E(y_{2n} - y_{2n-1})(Z_n - \theta) + \frac{a_n^2}{c_n^2} E(y_{2n} - y_{2n-1})^2. \tag{4}$$

It follows easily from Condition 1 that

$$-c_nK_2b_n \leqq E(y_{2n} - y_{2n-1})(Z_n - \theta) \leqq - c_nK_1b_n, \tag{5}$$

also from Conditions 1 and 2 we have

$$2M_1 + c_n^2K_1^2b_n \leqq E(y_{2n} - y_{2n-1})^2 \leqq 2M_2 + c_n^2K_2^2b_n. \tag{6}$$

Therefore from (4), (5), (6) and Condition 3 we get

$$\begin{aligned} b_n\left(1 - \frac{2K_2}{n^{1-\epsilon}}\right) + \frac{2M_1}{n^{1+2(\eta-\epsilon)}} &\leqq b_n\left(1 - \frac{2K_2}{n^{1-\epsilon}} + \frac{K_1^2}{n^{2(1-\epsilon)}}\right) + \frac{2M_1}{n^{1+2(\eta-\epsilon)}} \leqq b_{n+1} \\ &\leqq b_n\left(1 - \frac{2K_1}{n^{1-\epsilon}} + \frac{K_2^2}{n^{2(1-\epsilon)}}\right) + \frac{2M_2}{n^{1+2(\eta-\epsilon)}} \\ &= b_n\left(1 - \frac{1}{n^{1-\epsilon}}\left(2K_1 - \frac{K_2^2}{n^{1-\epsilon}}\right)\right) + \frac{2M_2}{n^{1+2(\eta-\epsilon)}}. \end{aligned} \tag{7}$$

For any $\delta > 0$, there exists an n_0 such that for $n > n_0$, $2K_1 - K_2^2 / n^{1-\epsilon} \geqq 2(K_1 - \delta)$. Therefore, by the second part of Chung's lemma, $\overline{\lim}_{n\to\infty} n^{2n-\epsilon}b_n \leqq M_2 / K_1 - \delta$. But since δ is arbitrary, the right side of (3) follows. The left side of (3) follows immediately from (7) and the first part of Chung's Lemma.

It is a corollary of Theorem 1 that Z_n converges stochastically to θ as $n \to \infty$. We remark that for stochastic convergence we need not impose the condition that $M_1 > 0$.

4. Convergence to the Normal Distribution. Let

$\beta^{(r)}(x) = \int_{-\infty}^{\infty} |y - M(x)|^r \, dH(y|x)$ for $r > 0$.

We shall need the following condition on the $\beta^{(r)}$'s.

CONDITION 4. *There exist real numbers* $M_1(r)$ *and* $M_2(r)$ *such that for all* x,

$$0 < M_1(r) \leqq \beta^{(r)}(x) \leqq M_2(r) < \infty, \qquad r = 1, 2, \cdots.$$

We shall continue to denote $M_1(2)$ and $M_2(2)$ by M_1 and M_2 respectively.

LEMMA 1. *If Conditions* 1, 3 *and* 4 *hold, then*

$$\begin{aligned} (2k-1)(2k-3)\cdots 3\cdot 1\left(\frac{M_1}{K_2}\right)^k &\leqq \underline{\lim}_{n\to\infty} n^{2k(\eta-\epsilon/2)}b_n^{(2k)} \\ \leqq \overline{\lim}_{n\to\infty} n^{2k(\eta-\epsilon/2)}b_n^{(2k)} &\leqq (2k-1)(2k-3)\cdots 3\cdot 1\left(\frac{M_2}{K_1}\right)^k, \quad k = 1, 2, \cdots, \end{aligned} \tag{8}$$

where $b_n^{(r)} = E(Z_n - \theta)^r$.

PROOF. The proof is by induction. By theorem 1, (8) is true when $k = 1$. From (1) and Condition 3 we have for any positive integer r,

$$\begin{aligned} b_{n+1}^{(r)} = b_n^{(r)} &+ \frac{r}{n^{1/2+\eta-\epsilon}} E(Z_n - \theta)^{r-1}(y_{2n} - y_{2n-1}) \\ &+ \frac{r(r-1)}{2n^{1+2(\eta-\epsilon)}} E(Z_n - \theta)^{r-2}(y_{2n} - y_{2n-1})^2 \\ &+ \sum_{j=3}^{r} \binom{r}{j} \frac{1}{n^{j/2+j(\eta-\epsilon)}} E(Z_n - \theta)^{r-j}(y_{2n} - y_{2n-1})^j. \end{aligned} \tag{9}$$

Imposing Conditions 1, 3 and 4 and utilizing the induction assumption, (9) yields for even r,

$$b_{n+1}^{(r)} \leqq b_n^{(r)}\left(1 - \frac{rK_1}{n^{1-\epsilon}} + \frac{r(r-1)}{2}\frac{K_2^2}{n^{2(1-\epsilon)}}\right) + \frac{r(r-1)\cdots 3\cdot 1}{n^{r(\eta-\epsilon/2)+(1-\epsilon)}}\left(\frac{M_2}{K_1}\right)^{(r-2)/2} + o(n^{-[r(\eta-\epsilon/2)+1-\epsilon]}).$$

The inequality on the right side of (8) follows, as before, from an application of Chung's lemma. The inequality on the left can be obtained in a similar manner, replacing M_2 by M_1 and K_1 by K_2, reversing the inequalities, and again applying Chung's lemma. This proves the lemma.

Condition 5. *$\sigma^2(x)$ is continuous at $x = \theta$.*

Condition 6. *There exist positive numbers δ, L, and C_0 such that for all c, $0 < c < C_0$,*

$$M(x + c) - M(x - c) = cK(x)(x - \theta)$$

where

$$-K' - W|x - \theta|^\delta \leqq K(x) \leqq -K' + W|x - \theta|^\delta \quad \text{for} \quad |x - \theta| \leqq L,$$
$$-K_2 \leqq K(x) \leqq -K_1 \quad \text{for} \quad |x - \theta| > L$$

and where W and K' are positive number such that $-K' + WL^\delta \leqq -K_1$ and $-K' - WL^\delta \geqq -K_2$.

Condition 6 is a strengthening of Condition 1 to the extent that locally at $x = \theta$, $M(x)$ is parabolic.

Lemma 2. *If Conditions 2, 3, 4, 5, 6 hold, then*

$$\lim_{n\to\infty} n^{2\eta-\epsilon} b_n = \frac{\sigma^2(\theta)}{K'} \tag{10}$$

Proof. From Condition 6 we have

$$c_n(-K'b_n - WE|Z_n - \theta|^{2+\delta}) \leqq E(y_{2n} - y_{2n-1})(Z_n - \theta) \leqq c_n(-K'b_n + \mathrm{W}\mathrm{E}|Z_n - \theta|^{2+\delta}).$$

From Lemma 1 and Lyapunov's inequality for n large enough and for some k

$$E\,|Z_n - \theta|^{2+\delta} \leqq (b_n^{(2k)})^{(2+\delta)/2k} \leqq \frac{R_k^{(2+\delta)/2k}\, b_n}{R_1 n^{(\eta-\epsilon/2)(1+\delta)}} \tag{11}$$

where R_k denotes the upper bound in (8) and R_1 denotes the lower bound in (8) for $k = 1$. Also,

$$E(y_{2n} - y_{2n-1})^2 = E(\sigma^2(Z_n + c_n) + \sigma^2(Z_n - c_n)) + c_n^2\gamma K_2^2 b_n$$

where $|\gamma| < 1$. But since Z_n converges stochastically to θ, $c_n \to 0$, and $\sigma^2(x)$ is bounded and continuous at θ, we have using (11)

$$b_{n+1} = b_n \left\{ 1 - \frac{2}{n^{1-\epsilon}} (K' - w_n) + \frac{K^2}{n^{2(1-\epsilon)}} \right\} + \frac{2\sigma^2(\theta) + d_n}{n^{1+2(\eta-\epsilon)}}, \tag{12}$$

where both w_n and d_n tend to zero as $n \to \infty$. Both parts of Chung's lemma can be applied to (12). Therefore, (10) follows.

LEMMA 3. *If Conditions* 2, 3, 4, 5, 6 *hold, then* $\lim_{n\to\infty} n^{\eta-\epsilon/2} b_n^{(1)} \to 0$.

PROOF. It follows from (1) and Condition 6 that

$$\begin{aligned} b_n^{(1)} \left(1 - \frac{K'}{n^{1-\epsilon}}\right) - \frac{W}{n^{1-\epsilon}} E\,|Z_n - \theta|^{1+\delta} &\leqq b_{n+1}^{(1)} \\ &\leqq b_n^{(1)} \left(1 - \frac{K'}{n^{1-\epsilon}}\right) + \frac{W}{n^{1-\epsilon}} E\,|Z_n - \theta|^{1+\delta}. \end{aligned} \tag{13}$$

But by Lemma 1 and Lyapunov's inequality, there is a constant $R > 0$ dependent on δ such that

$$E\,|Z_n - \theta|^{1+\delta} \leqq \frac{R}{n^{(1+\delta)(\eta-\epsilon/2)}}. \tag{14}$$

From (13) and (14) and the remark following Chung's lemma it follows that

$$\overline{\lim_{n\to\infty}}\, n^{\eta-\epsilon/2} b_n^{(1)} \leqq 0. \tag{15}$$

Also, it can be shown in the same way that

$$\overline{\lim_{n\to\infty}}\, n^{\eta-\epsilon/2} (-b_n^{(1)}) \leqq 0. \tag{16}$$

Lemma 3 follows from (15) and (16).

THEOREM 2. *If Conditions* 2, 3, 4, 5, 6 *hold, then* $n^{\eta-\epsilon/2}(Z_n - \theta)$ *converges in distribution to a normal distribution with mean zero and variance* $\sigma^2(\theta)/K'$.

PROOF. By using induction, it follows from lemmas 1, 2 and 3, and Lyapunov's inequality that

$$\begin{aligned} \lim_{n\to\infty} n^{r(\eta-\epsilon/2)}\, b_n^{(r)} &= \left(\frac{\sigma^2(\theta)}{K'}\right)^{r/2} (r-1)(r-3) \cdots 3\cdot 1 \text{ for } r \text{ even, and} \\ &= 0 \text{ for } r \text{ odd} \end{aligned} \tag{17}$$

The method of induction is similar to that used in proving Lemma 1. We shall omit the details. However, (17) indicates that the moments of $n^{\eta-\epsilon/2}(Z_n - \theta)$ converge to the moments of the indicated normal distribution. The result follows from the well-known theorem on the convergence of moments.

REFERENCES

[1] J. R. BLUM, "Approximation Methods which converge with probability one," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 382–386.

[2] K. L. CHUNG, "On a stochastic approximation method," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 463–483.

[3] J. KIEFER AND J. WOLFOWITZ, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 462–466.

[4] H. ROBBINS AND S. MONRO, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400–407.