THE PROBABILITY IN THE TAIL OF A DISTRIBUTION

By Melvin L. Katz[1]

University of Chicago

Introduction. Let $X_1, X_2, \cdots$ denote a sequence of independent and identically distributed random variables. If the first moment of X_k is finite then it is well known that $n^{-1}\sum_{k=1}^{n} X_k \to \mu$ with probability one where $\mu = EX_k$. For $\epsilon > 0$ and $n = 1, 2, \cdots$ define $p_n(\epsilon) = P\{|\sum_{k=1}^{n} X_k - n\mu| > n\epsilon\}$; then $p_n(\epsilon) \to 0$ as $n \to \infty$ and it is the purpose of this paper to study the rate of convergence to zero of the p_n's. Recently several papers have considered this problem, for example [1], [3], and other papers have considered the same problem for sequences of random variables which are not necessarily independent and identically distributed, for example [5]. In all the papers investigating the rate of convergence of the p_n's it is assumed that the moment generating function of the X_k's is finite in some interval and then it is shown that $p_n(\epsilon) = O(\rho^n(\epsilon))$ where $\rho(\epsilon) < 1$. In [2] it is shown that the existence of the moment generating function of the X_k's is both necessary and sufficient for $p_n(\epsilon) = O(\rho^n(\epsilon))$ in the case of independent and identically distributed random variables and necessary for many other classes of processes.

In the case that the moment generating functions of the random variables under consideration do not exist the problem of the rate of convergence of the p_n's to zero does not seem to have been so well investigated and it is the purpose of this paper to consider this problem for independent and identically distributed random variables. The rate of convergence of $p_n(\epsilon)$ to zero is determined in Theorem 1 if $E|X_k|^t < \infty$ for some $t \geqq 1$. In Theorems 2 and 3 results are obtained on the analogous problems if $E|X_k| = +\infty$ and $E|X_k|^t < \infty$ for $t < 1$, or if moments higher than the first exist and the sums $\sum_{k=1}^{n} (X_k - \mu)$ are normed by n^α where $\alpha > \frac{1}{2}$.

Theorem 1 of this paper has been proved in the case $t = 1$ by Spitzer [1] and the case $t = 2$ by Erdös [4]. The proofs of this paper rely heavily on the methods of [4].

Theorems. Let $\{X_k : k = 1, 2, \cdots\}$ be a sequence of independent identically distributed random variables.

Theorem 1. *Let* $t \geqq 1$. *Then* $E|X_k|^t < \infty$ *and* $EX_k = \mu$ *if and only if*

$$\sum_{n=1}^{\infty} n^{t-2} P\left\{\left|\sum_{k=1}^{n} X_k - n\mu\right| > n\epsilon\right\} < \infty$$

for all $\epsilon > 0$. We shall defer the proof of Theorem 1 until Theorems 2 and 3 have been proved.

Received March 8, 1962.

[1] Research carried out in the Department of Statistics, University of Chicago with the partial support of the Office of Naval Research, Contract Nonr-2121(09).

THEOREM 2. *Let* $t > 0$ *and* $r > 1$. (a) *If* $t \geqq 1$ *and* $\frac{1}{2} < r/t \leqq 1$, *then* $E|X_k|^t < \infty$ *and* $EX_k = \mu$ *imply* $\sum_{n=1}^{\infty} n^{r-2} P\{|\sum_{k=1}^{n} X_k - n\mu| > n^{r/t}\epsilon\} < \infty$ *for all* $\epsilon > 0$. (b) *If* $t \geqq 1$ *and* $r/t > 1$, *then* $E|X_k|^t < \infty$ *implies* $\sum_{n=1}^{\infty} n^{r-2} P\{|\sum_{k=1}^{n} X_k| > n^{r/t}\epsilon\} < \infty$ *for all* $\epsilon > 0$. (c) *If* $t < 1$ *and* $r/t > 1$, *then* $E|X_k|^t < \infty$ *implies* $\sum_{n=1}^{\infty} n^{r-2} P\{|\sum_{k=1}^{n} X_k| > n^{r/t}\epsilon\} < \infty$ *for all* $\epsilon > 0$.

Thus, for example, if $EX_k = \mu$ and $E|X_k|^{10} < \infty$ Theorem 2 asserts that $p_n = o(1/n^8)$, $P\{|\sum_{k=1}^{n} X_k - n\mu| > n^{4/5}\epsilon\} = o(1/n^6)$ and $P\{|\sum_{k=1}^{n} X_k| > n^2\epsilon\} = o(1/n^{18})$.

PROOF OF THEOREM 2. To prove this theorem we may assume with no loss of generality that $\epsilon = 1$ and, if it exists, that $EX_k = 0$. Following the methods of [4] we define

$$A_n = \left\{\left|\sum_{k=1}^{n} X_k\right| > n^{r/t}\right\}$$

and

$$a_i = P\{|X_k| > 2^{ir/t}\}.$$

Then it is easy to see that $\sum_{i=0}^{\infty} 2^{ir} a_i < \infty$ is equivalent to the condition $E|X_k|^t < \infty$. Let $2^i \leqq n < 2^{i+1}$ and define

$$A_n^{(1)} = \{|X_k| > 2^{(i-2)r/t} \quad \text{for at least one} \quad k \leqq n\}$$

$$A_n^{(2)} = \{|X_{k_1}| > n^{\gamma r/t}, |X_{k_2}| > n^{\gamma r/t} \quad \text{for at least two} \quad k\text{'s} \leqq n\},$$

where γ is chosen so that $[(r + 1)/2r] < \gamma < 1$, $(1 - \gamma r) < 0$ and $[1 - (2\gamma r/t)] < 0$. Such a choice is possible by hypothesis.

$$A_n^{(3)} = \left\{\left|\sum_{k=1}^{n}{}' X_k\right| > 2^{(i-2)r/t}\right\},$$

where $\sum'$ denotes the sum omitting those k with $|X_k| > n^{\gamma r/t}$. Since $r/t > \frac{1}{2}$ it follows that

$$A_n \subset A_n^{(1)} \cup A_n^{(2)} \cup A_n^{(3)}. \tag{1}$$

Therefore to prove the theorem it will be sufficient to prove that

$$\sum_{n=1}^{\infty} n^{r-2} P(A_n^{(j)}) < \infty \qquad \text{for } j = 1, 2, 3.$$

Since we have assumed $EX_k = 0$ it is necessary in part (b) to prove that

$$\sum_{n=1}^{\infty} n^{r-2} P\{|\sum_{k=1}^{n} X_k - n\mu| > n^{r/t}\epsilon\} < \infty \qquad \text{for all } \epsilon > 0.$$

However, since $r/t > 1$ this is equivalent to showing that

$$\sum_{n=1}^{\infty} n^{r-2} P\{|\sum_{k=1}^{n} X_k| > n^{r/t}\epsilon\} < \infty \qquad \text{for all } \epsilon > 0$$

and thus it is enough to show $\sum_{n=1}^{\infty} n^{r-2} P(A_n^{(j)}) < \infty$ for $j = 1, 2, 3$. We begin by demonstrating that $\sum_{n=1}^{\infty} n^{r-2}P(A_n^{(1)}) < \infty$. We have that

$$P(A_n^{(1)}) = P\{|X_k| > 2^{(i-2)r/t} \quad \text{for some} \quad k \leqq n\} \tag{2}$$
$$\leqq nP\{|X_k| > 2^{(i-2)r/t}\} = na_{i-2} \leqq 2^{i+1}a_{i-2}.$$

Thus

$$\sum_{n=1}^{\infty} n^{r-2}P(A_n^{(1)}) = \sum_{i=0}^{\infty} \sum_{2^i \leqq n < 2^{i+1}} n^{r-2}\cdot 2^{i+1}a_{i-2} < \sum_{i=0}^{\infty} 2^{(2i+1)+(i+1)(r-2)}a_{i-2} \tag{3}$$
$$= \text{constant} + \sum_{i=0}^{\infty} 2^{ri+3r-1}a_i < \infty.$$

To show $\sum_{n=1}^{\infty} n^{r-2}P(A_n^{(2)}) < \infty$ we proceed as follows.

$$P(A_n^{(2)}) = P\{|X_{k_1}| > n^{\gamma r/t}, |X_{k_2}| > n^{\gamma r/t} \quad \text{for at least two} \quad k\text{'s} \leqq n\}$$
$$\leqq \sum_{1 \leqq k_1 < k_2 \leqq n} P\{|X_{k_1}| > n^{\gamma r/t}, |X_{k_2}| > n^{\gamma r/t}\} \tag{4}$$
$$\leqq n^2P^2\{|X_k| > n^{\gamma r/t}\}.$$

Now $P\{|X| > n^{\gamma r/t}\} \leqq (E|X|^t/n^{\gamma r}) = c/n^{\gamma r}$ and therefore

$$P(A_n^{(2)}) \leqq (c\, n^2/n^{2\gamma r}) = c\, n^{2(1-\gamma r)}. \tag{5}$$

Throughout the proof c will denote all constants. Thus even in one set of inequalities we shall use c to denote two different constants. Thus it follows that $\sum_{n=1}^{\infty} n^{r-2}P(A_n^{(2)}) \leqq c \sum_{n=1}^{\infty} n^{r(1-2\gamma)}$. However, γ has been chosen so that $r(1 - 2\gamma) < -1$, since $[(r + 1)/2r] < \gamma$, and thus $\sum_{n=1}^{\infty} n^{r-2}P(A_n^{(2)}) < \infty$.

It remains to check the convergence properties of $\sum_{n=1}^{\infty} n^{r-2}P(A_n^{(3)})$. We begin by proving convergence in the case $t < 1$ and $r/t > 1$. Let $\delta > 0$ be such that $t + 2\delta = 1$ and define

$$X_k^+ = \begin{cases} X_k & \text{if} \quad |X_k| \leqq n^{\gamma r/t} \\ 0 & \text{otherwise.} \end{cases}$$

Then $P(A_n^{(3)}) = P\{|\sum_{k=1}^{\prime n} X_k| > 2^{(i-2)r/t}\} = P\{|\sum_{k=1}^{n} X_k^+| > 2^{(i-2)r/t}\}$. Further $2^i \leqq n < 2^{i+1} \Rightarrow n/8 < 2^{(i-2)}$ and therefore

$$P(A_n^{(3)}) \leqq P\left\{\left|\sum_{k=1}^{n} X_k^+\right| > n^{r/t}/8^{r/t}\right\}$$
$$\leqq c\, E\left|\sum_{k=1}^{n} X_k^+\right|^{t+\delta} \Big/ n^{r+\delta r/t}$$
$$\leqq c \sum_{k=1}^{n} E|X_k^+|^{t+\delta}/n^{r+\delta r/t} \tag{6}$$
$$\leqq c\, n^{\gamma r\delta/t} \sum_{k=1}^{n} E|X_k^+|^t/n^{r+r\delta/t}.$$

Therefore $n^{r-2}P(A_n^{(3)}) \leqq c\, n^{-1-(r/t)\delta(1-\gamma)}$ and since γ has been chosen so that $(1-\gamma) > 0$ it follows that if $t < 1$ and $r/t > \frac{1}{2}$ then $\sum_{n=1}^{\infty} n^{r-2}P(A_n^{(3)}) < \infty$. This completes the proof of part (c) of the theorem and we proceed to prove that if $t \geqq 1$ and $\frac{1}{2} < r/t \leqq 1$ then $\sum_{n=1}^{\infty} n^{r-2}P(A_n^{(3)}) < \infty$ i.e., part (a) will be proved. Recall that we have chosen $\mu = 0$.

We now take j to be the smallest integer $\geqq t$, and we let M be a positive integer to be determined later. Define $\alpha = EX_k^+$ and note that $\alpha \to 0$ as $n \to \infty$ since μ has been chosen $= 0$. Define $Y_k = X_k^+ - \alpha$ and note that $EY_k = 0$. We will find a bound for $P(A_n^{(3)})$ by using Markov's inequality and this requires that we find a bound for $E|\sum_{k=1}^{n} Y_k|^{2Mj}$

$$(7) \quad E\left|\sum_{k=1}^{n} Y_k\right|^{2Mj} = \sum_{k=1}^{n} EY_k^{2Mj} + \cdots + c \sum_{k_1<k_2<\cdots<k_\tau} EY_{k_1}^2 EY_{k_2}^2 \cdots EY_{k_\tau}^2,$$

where $\tau \leqq Mj$, and we bound $E|\sum_{k=1}^{n} Y_k|^{2Mj}$ by bounding each sum on the right hand side of (7).

$$(8) \qquad EY_k^{2Mj} = E|Y_k|^{2Mj-t}|Y_k|^t \leqq (n^{\gamma r/t} + |\alpha|)^{2Mj-t}E|X_k| + |\alpha|)^t$$

and therefore

$$(9) \qquad \textstyle\sum_{k=1}^{n} EY_k^{2Mj} \leqq c\, n^{\{(2Mj\gamma r/t)-\gamma r+1\}}.$$

Now consider any other sum on the right hand side of (7) where at least one of the exponents of the Y_{k_i}'s is greater than t. Suppose that for one of these sums exactly q of the exponents in each summand exceed t and l of the exponents are less than or equal to t. Then this sum is bounded by

$$(10) \qquad \begin{aligned} &c\, n^{1+(\gamma r/t)\{(t/\gamma r)[q+l-1]+(d1-t)+\cdots+(d_q-t)\}} \\ &\leqq c\, n^{1+(\gamma r/t)\{(t/\gamma r)[q+l-1]+(2Mj-2l)-qt\}} = c\, n^{(2Mj\gamma r/t)+q(1-\gamma r)+l(1-(2\gamma r/t))} \end{aligned}$$

where $d_1, \cdots, d_q$ are the exponents in each summand that exceed t. Now γ has been chosen so that $(1-\gamma r) < 0$ and $[1-(2\gamma r/t)] < 0$ and thus this bound is maximized when $q = 1$ and $l = 0$. That is all sums that have summands where at least one exponent of a Y_{k_i} is $> t$ are bounded by

$$(11) \qquad c\, n^{\{(2Mj\gamma r/t)-\gamma r+1\}}.$$

If all the exponents of the Y_{k_i} for a particular sum on the right hand side of (7) are $\leqq t$, then a bound for such a sum is given by $c\, n^{Mj}$. However, if M is sufficiently large

$$(12) \qquad (2Mj\gamma r/t) - \gamma r + 1 > Mj.$$

Choose an M sufficiently large so that (12) holds. Thus

$$(13) \qquad E\left|\sum_{k=1}^{n} Y_k\right|^{2Mj} \leqq c\, n^{\{(2Mj\gamma r/t)-\gamma r+1\}}.$$

If $r/t = 1$, then

$$(14)\qquad \begin{aligned} P(A_n^{(3)}) &= P\left\{\left|\sum_{k=1}^{n}{}' X_k\right| > 2^{(i-2)}\right\} \\ &= P\left\{\left|\sum_{k=1}^{n} X_k^+\right| > 2^{(i-2)}\right\} \\ &= P\left\{\left|\sum_{k=1}^{n} Y_k + n\alpha\right| > 2^{(i-2)}\right\}. \end{aligned}$$

Now choose n so large that $\alpha < \frac{1}{16}$ and since $n/8 < 2^{i-2}$ it follows that

$$(15)\qquad P(A_n^{(3)}) \leqq P\left\{\left|\sum_{k=1}^{n} Y_k\right| > \frac{n}{16}\right\} \leqq \left\{c\, E\left(\sum_{k=1}^{n} Y_k\right)^{2Mj} \Big/ n^{2Mj}\right\}.$$

Therefore, for n sufficiently large

$$(16)\qquad n^{r-2}P(A_n^{(3)}) \leqq c\, n^{2Mj(\gamma-1)-r(\gamma-1)-1}.$$

However, $(\gamma - 1) < 0$ and $2Mj > r$ and therefore

$$(17)\qquad \sum_{n=1}^{\infty} n^{r-2}P(A_n^{(3)}) < \infty.$$

If $\frac{1}{2} < r/t < 1$ we first note that it follows from integration by parts and the existence of EX_k that

$$(18)\qquad n^{1-r/t} \int_{|x|>n^{\gamma r/t}} xP\,(dx) \to 0 \qquad \text{as } n \to \infty.$$

Thus

$$(19)\qquad \begin{aligned} P(A_n^{(3)}) &\leqq P\left\{\left|\sum_{k=1}^{n} Y_k + n\alpha\right| > c\, n^{r/t}\right\} \\ &\leqq P\left\{\left|\sum_{k=1}^{n} Y_k\right| > c\, n^{r/t}(1 - c\, n^{1-r/t}\alpha)\right\}. \end{aligned}$$

However,

$$\left|\int_{|x|>n^{\gamma r/t}} xdP\right| = \left|\int_{-\infty}^{\infty} xdP - \alpha\right| = |\alpha|$$

and thus $n^{1-r/t}\alpha \to 0$ as $n \to \infty$ since $EX_k = 0$. Therefore for large n

$$(20)\qquad P(A_n^{(3)}) \leqq P\left\{\left|\sum_{k=1}^{n} Y_k\right| \geqq c\, n^{r/t}\right\} \leqq \left\{c\, E\left(\sum_{k=1}^{n} Y_k\right)^{2Mj} \Big/ n^{2Mjr/t}\right\}$$

and as before we have that $\sum_{n=1}^{\infty} n^{r-2}P(A_n^{(3)}) < \infty$. This completes the proof of part (a). It remains to prove part (b).

Recall that for part (b) of the theorem we assume that $t \geqq 1$ and $r/t > 1$.

Therefore for sufficiently large n

$$P(A_n^{(3)}) = P\left\{\left|\sum_{k=1}^{n}{}' X_k\right| > 2^{(i-2)r/t}\right\}$$

$$(21) \qquad = P\left\{\left|\sum_{k=1}^{n} Y_k\right| > c\, n^{r/t} - n\alpha\right\}$$

$$\leqq P\left\{\left|\sum_{k=1}^{n} Y_k\right| > c\, n^{r/t}\right\}.$$

Note that under the hypotheses of part (b) α is not necessarily converging to zero as $n \to \infty$ but since $r/t > 1$ this is immaterial. Finally we apply Markov's inequality to (21) and it follows from (13) that $\sum n^{r-2}P(A_n^{(3)}) < \infty$.

A partial converse to Theorem 2 is provided by Theorem 3.

THEOREM 3. *Let* $t > 0$ *and* $r \geqq 2$. (a) *If* $t \geqq 1$ *and* $\frac{1}{2} < r/t \leqq 1$, *then* $\sum_{n=1}^{\infty} n^{r-2}P\{|\sum_{k=1}^{n} X_k - n\mu| > n^{r/t}\epsilon\} < \infty$ *for all* $\epsilon > 0$ *implies that* $E|X_k|^t < \infty$ *and* $EX_k = \mu$. (b) *If* $t \geqq 1$ *and* $r/t > 1$, *then* $\sum_{n=1}^{\infty} n^{r-2}P\{|\sum_{k=1}^{n} X_k| > n^{r/t}\epsilon\} < \infty$ *for all* $\epsilon > 0$ *implies that* $E|X_k|^t < \infty$. (c) *If* $t < 1$ *and* $r/t > 2$, *then* $\sum_{n=1}^{\infty} n^{r-2}P\{|\sum_{k=1}^{n} X_k| > n^{r/t}\epsilon\} < \infty$ *for all* $\epsilon > 0$ *implies* $E|X_k|^t < \infty$.

PROOF. To prove that in part (a) $EX_k = \mu$ we just observe that the hypotheses of the theorem imply that $P\{\lim_{n} n^{-1}\sum_{k=1}^{n} (X_k - \mu) = 0\} = 1$ and thus $EX_k = \mu$. To check that $E|X_k|^t < \infty$ in part (a) and in parts (b) and (c) we follow the method of [4] exactly. If we set $Y_k = |X_k|^{t/r}$ it follows as in [4] that $P\{|\sum_{k=1}^{n} X_k| > n^{r/t}\} \geqq nP\{|Y_n| > c\,n\}$. Thus $\sum_{n=1}^{\infty} n^{r-2}P\{|\sum_{k=1}^{n} X_k| > n^{r/t}\} < \infty$ implies that $\sum_{n=1}^{\infty} n^{r-1}P\{|Y_n| > c\,n\} < \infty$ and therefore $E|Y_n|^r < \infty$. Since $E|Y_n|^r = E|X_n|^t$ the proof is complete.

We proceed now to the proof of Theorem 1.

PROOF OF THEOREM 1. It follows from Theorem 2, for $t > 1$, and from Spitzer [7], for $t = 1$, that if $EX_k = \mu$ and $E|X_k|^t < \infty$ then $\sum_{n=1}^{\infty} n^{t-2}P\{|\sum_{k=1}^{n} X_k - n\mu| > n\epsilon\} < \infty$ for all $\epsilon > 0$. If $t \geqq 2$ it follows from Theorem 3 that if $\sum_{n=1}^{\infty} n^{t-2}P\{|\sum_{k=1}^{n} X_k - n\mu| > n\epsilon\} < \infty$ for all $\epsilon > 0$ then $EX_k = \mu$ and $E|X_k|^t < \infty$. For $t = 1$ Spitzer has proved this fact and therefore we must only show this for $1 < t < 2$.

If $\sum_{n=1}^{\infty} n^{t-2}P\{|\sum_{k=1}^{n} X_k - n\mu| > n\epsilon\} < \infty$ for all $\epsilon > 0$ and $1 < t < 2$ then $\sum_{n=1}^{\infty} n^{-1}P\{|\sum_{k=1}^{n} X_k - n\mu| > n\epsilon\} < \infty$ for all $\epsilon > 0$ and from Spitzer's result it follows that $EX_k = \mu$. To show that $E|X_k|^t < \infty$ we may again employ the method of [4] provided we can show that $nP\{|X_n| > 2n\} \to 0$ as $n \to \infty$ and that there exists a constant $\tau > 0$, independent of k and n, for n sufficiently large such that

$$P\{|\sum_{\substack{l=1\\k\neq l}}^{n} (X_l - \mu)| < n\} \geqq \tau.$$

However, since we know that EX_k exists it follows from the moments lemma [6, pg. 242] that $nP\{|X_n| > 2n\} \to 0$ and from the Law of Large Numbers that

τ exists. Therefore the results of [4] are applicable and we obtain that $P\{|\sum_{k=1}^{n} X_k| > n\} \geqq nP\{|X_n| > cn\}$. ($\mu$ has been set $= 0$.) Thus $\sum_{n=1}^{\infty} n^{t-1} P\{|X_n| > c\,n\} < \infty$ and hence $E|X_k|^t < \infty$.

REFERENCES

[1] BAHADUR, R. R. and RANGA RAO, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[2] BAUM, LEONARD E., KATZ, MELVIN and READ, ROBERT R. (1962). Exponential convergence rates for the law of large numbers. *Trans. Amer. Math. Soc.* **102** 187–199.

[3] BLACKWELL, DAVID and HODGES, JR., J. L. (1959). The probability in the extreme tail of a convolution. *Ann. Math. Statist.* **30** 1113–1120.

[4] ERDÖS, P. (1949). On a theorem of Hsu and Robbins. *Ann. Math. Statist.* **20** 286–291.

[5] KOOPMANS, L. H. (1961). An exponential bound on the strong law of large numbers for linear stochastic processes with absolutely convergent coefficients. *Ann. Math. Statist.* **32** 583–586.

[6] LOÈVE, M. (1960). *Probability Theory.* Van Nostrand, New York.

[7] SPITZER, FRANK (1956). A combinatorial lemma and its application to probability theory. *Trans. Amer. Math. Soc.* **82** 323–339.