CLASSICAL TAUBERIAN THEOREMS

By D. D. Kosambi

(Poona)

A MODERN statement of the general Tauberian theorem runs as follows: "If G is locally compact Abelian group, then every proper closed ideal of $L^1(G)$ is included in a regular maximal ideal ".[1] Neither this, nor the work of N. Wiener[2] from which it was generalized, furnishes any motivation for the classical forms. None of these approaches, including that of S. Bochner,[3] are of any help in finding the Tauberian convergence conditions associated with a given regular method of summability. This note points out the existence of a unified method for deriving the major classical Tauberian theorems, *defined* as those obeying conditions (1) and (2) below.

In what follows, the bounded sequences $\{s_n\}$ are the partial sums of an infinite series of real terms Σa_n, so that $a_n = s_n - s_{n-1}$. Summability means replacement of $\{s_n\}$ by linear transforms:

$$s'(t) = \sum_{n=1}^{\infty} P_n(t)\, s_n; \quad P_n(t) \text{ real with real } t \geq 0. \tag{1}$$

If $s'(t) \to s$ as t tends to some fixed critical value t^*, (which we shall eventually take as $t^* = \infty$) then the series Σa_n is said to be summable to s by that method of summability. The method is *regular* if any series Σa_n is summable to the value s whenever it converges to s. The conditions necessary and sufficient thereto are that $\Sigma P_n(t) \to 1$, and $P_n(t) \to 0$ for each n as $t \to t^*$.

It is noteworthy that the commonly used methods[4] of summability obey more stringent conditions besides those of regularity, namely:—

$$\left.\begin{matrix}(a)\ P_n(t) \geq 0 \\ (b)\ \Sigma P_n(t) = 1\end{matrix}\right\} \text{ for all } t; \quad (c)\ P_n(t) \to 0 \text{ as } t \to t^*. \tag{2}$$

These make $\{P_n(t)\}$ a probability distribution for each value of t. That is, $P_n(t)$ may be treated as the probability with which a stochastic variable X_t assumes the integral value n. The sum in (1) is then the *expectation* (mean value) of the function $s_n = s(X)$. This suggests the further use of probability methods. The particular method followed here is to derive a continuous limiting distribution from $P_n(t)$ as $t \to t^*$. If then s_n can simultaneously be made to tend to a continuous limiting function, the problem of convergence reduces to the form of that limit function. The method followed means, in general, the use of an increasingly larger scale for the real variable which initially assumed the integral values n. Though illustrated for infinite series, the procedure is also valid for integrals. The conditions (2) are also related to certain Markov processes, but that will not be considered here.

Summability procedures of the type given in (1) and (2) measure the limiting density of the sequence. For convergent sequences, the density is unity at the value s, and zero everywhere else, in the limit. This means, in probability terms, that the (limiting) dispersion or variance, the expectation of $(s_n - s)^2$ must vanish. That is equivalent to the statement that the expectation of s_n^2 is s^2, in the limit. Similarly for the higher moments. This gives us:

LEMMA 1: *For a convergent* $\{s_n\}$, $a_n = (s_n - s_{n-1}) \rightarrow 0$, *and* $\{s_n^k\}$ *has the sum* s^k *by any regular method of summability.*

These necessary conditions are obvious, and hold for any regular method, whether (2) is fully obeyed or not. The restriction on the k-th moment does not, without convergence, imply $a_n \rightarrow 0$. This may be seen from the example: $s_n = 1$ when $n = m^2$ and $s_n = 0$ otherwise, which is summable $(C, 1)$ to zero, with all its moments, though a_n tends to no limit.

If the limiting probability distribution and the limiting function derived from $\{s_n\}$ both exist, the vanishing of the dispersion becomes a strong condition for summable sequences:

LEMMA 2: *If a function* $f(x)$ *of a stochastic variable has an expectation, and if its variance (dispersion) vanishes,* i.e.,

$$\int f(x)\, dF = \bar{f}; \qquad \int \{f(x) - \bar{f}\}^2 dF = 0 \tag{3}$$

then $f(x) = \bar{f}$ *at each jump in* F *and almost everywhere on the range where* F *has a non-vanishing derivative.*

The proof is elementary, as $dF \geq 0$ for every probability distribution, the integrals being taken in the Lebesgue-Stieltjes sense. It will be seen that $dF > 0$ is essential over the entire range upon which the stochastic variable is mapped. This means that (2) cannot be generalized to cases where an arbitrary number of $P_n(t)$ may be zero or negative. Even if an infinite subsequence of $P_n(t)$ vanishes identically in t, nothing will be said about the behaviour of the corresponding subsequence of $\{s_n\}$, which would have to be deleted for the Tauberian theorem to have any meaning.

I use this condition on the second moment solely as a temporary prop. It will be shown that its purpose is served by the summability itself.

2. Let an analytic function $\phi(t)$, real over the real axis, be defined by the power series:

$$\phi(t) = b_0 + b_1 t + b_2 t^2 + \ldots + b_n t^n + \ldots; b_n > 0 \text{ for all large } n. \quad (4)$$

Then $P_n(t) = b_n t^n / \phi(t)$ can be used in (1) and (2) for all $t > 0$ throughout the interval of convergence. The critical value t^* would be the right-hand closure of the interval, which must be open to the right. Borel summability is given by $\phi(t) = e^t$, $t^* = \infty$. For Abel summability, take $\phi = 1/(1-t)$, $t^* = 1 - 0$, and so on.

The statistical properties of such functions as obey (4) have been discussed elsewhere,[5] and are quite easy to derive. In particular, the expectation $\mu(t)$ and variance $\sigma^2(t)$ of the stochastic variable X which takes on the value n with probability $b_n t^n / \phi(t)$ are:

$$\mu = \frac{1}{\phi} \sum n b_n t^n = \frac{t\phi'}{\phi} \; ; \; \sigma^2 = \frac{1}{\phi} \sum (n - \mu)^2 b_n t^n = \frac{d^2 \log \phi}{d \log t^2} \quad (5)$$

To apply these results, we have first to discuss conditions under which a limiting function $f(x)$ may be obtained from $\{s_n\}$. We start with $f(x, t_0)$, for some t_0 short of the critical value, defined as follows. The function $f(n, t_0) = s_n$; for $n - 1 < x \leq n$ the graph of $f(x, t_0)$ is given by the straight line joining the ordinates s_{n-1}, s_n. This amounts to replacement of s_n in the summation by the moving average $(s_{n+1} + s_n)/2$ if one should extend the value $P_n(t)$ to hold over the interval $(n, n+1)$. It may be possible to construct cases where the new sequences are not summable at all, let alone summable to the previous value; but this is immaterial to our arguments, because such pathological sequences will not converge in any case, and we are trying to derive conditions under which the summable sequence converges. So, it may be assumed that our sequences permit this operation. The $f(x, t_0)$ thus defined is bounded and continuous. Then we change the x-scale, which is allowed to tend to ∞ as $t \to t^*$. This gives an indexed (directed) set $f(x, t)$. The conditions for the existence of a continuous limiting function are given by:

LEMMA 3: *The indexed infinite set $f(x, t)$ has a subset which tends to a continuous limiting function $f(x)$ with modified difference quotient for any x-neighbourhood, if over that neighbourhood the $f(x, t)$ and its* *difference-quotient in x are uniformly bounded with the same bounds for all values of the index t.*

The proof follows from the classic theorem of Ascoli[6] on uniformly bounded smooth sequences of continuous functions with uniformly bounded difference quotients. For our purpose and special method of construction, the limit of the sub-set is enough, because only large values of the index n will appear, as the scale is increased, in any x-neighbourhood. All functions $f(x, t)$ of the set have the same bounds as $\{s_n\}$, which was a bounded sequence by hypothesis. The difference-quotient for $f(x, t_0)$ is either $s_n - s_{n-1}$, or lies between two adjacent values a_{n-1}, a_n in the neighbourhood of any point x. For change of scale by a factor λ, this difference-quotient remains bounded if and only if λa_n is bounded. It will be shown that summability serves the purpose of preventing more than one limit from existing, while the Tauberian condition $\lambda a_n = 0(1)$ guarantees that every infinite sub-sequence has at least one smooth limiting function. This is the substance of the simpler classical Tauberian theorems, excluding the 'high indices' theorem.

3. We proceed to find the actual transformations for important cases. For Borel summability, $\phi(t) = \exp t$, which is the Poisson distribution. The largest term $t^n/n!$ of the expansion of $\phi(t)$ is when n is the integer nearest to t. Moreover, $\mu = \sigma^2 = t$ for the mean value and variance. The proper transformation is therefore $x = (n - t)/\sqrt{t}$, where n is written for brevity for the original, continuous variable whose integral values furnish the index for a_n and s_n. This gives a normal (Gaussian) distribution in the limit:

$$dF = \left(\frac{1}{\sqrt{2\pi}}\right) e^{-x^2/2}\, dx. \tag{6}$$

This is well known from the properties of the Poisson distribution. The same thing happens when $\phi(t)$ is an integral function of finite order. The transformation is:

$$x = \frac{(n - \mu)}{\sigma}, \tag{7}$$

and the same limiting distribution is obtained as in (6). This is easily seen from:

LEMMA 4: *An integral function whose Taylor series* (4) *has only positive real coefficients is of finite order* ρ *if and only if* $\overline{\lim.}\ \sigma^2/\mu = \rho$.

Proof.—From the Hadamard factorization theorem for integral functions, it is easily shown that $\overline{\lim}.\ \sigma^2/\mu = \rho$, the order of the integral function. For the converse, we have by definition:

$$\frac{\sigma^2}{\mu} \equiv 1 + \frac{tf''}{f'} - \frac{tf'}{f} > 0.$$

If the superior limit of σ^2/μ is $\rho \geq 0$, integration gives us $a.\exp.(b.t^{\rho+\epsilon})$ for every $\epsilon > 0$ as a bounding function for f from above. This makes f an integral function of order ρ, by definition. It need not be pointed out that polynomials are excluded from the discussion by the condition that all Taylor coefficients must (eventually) be positive.

The scale factor is the standard deviation σ. For any fixed value of x, $n = \sigma x + \mu$, and the dominating term is clearly μ. As a function of t, μ is asymptotic to n for that x-neighbourhood. Thus the scale is essentially proportional to $\sqrt{n}$, so that the difference quotient remains bounded with $\sqrt{n}.a_n$. This leads to:

THEOREM 1: *For $\phi(t)$ in* (4) *an entire function of finite order and $P_n = b_n t^n/\phi(t)$, if $\{s_n\}$ and $\{s_n^2\}$ are summable respectively to s and s^2 with $\sqrt{n}.a_n = 0(1)$, then $\Sigma\, a_n$ converges to s.*

Proof.—An integrable limiting function $f(x)$ formed from the graphs of sequences exists by lemma 3. This must be constant almost everywhere by lemma 2. As it is continuous, it is identically a constant. Lemma 3, like the original Ascoli theorem, clearly applies to *every* infinite sub-set of $f(x, t)$. Therefore, if every infinite sub-set that converges to some limit function be struck off, at most a finite number of functions of the directed set can survive the deletion and they may be ignored. All the limit functions will have the same bounds, and all be continuous, with the same bound for the difference-quotient. However, the summability condition then says that the mean value $\bar{f}$ is the same for all, being the sum. Further, the vanishing dispersion, as in lemma 2, makes them all coincident, and equal to the constant $\bar{f}$. Therefore, the set $f(x, t)$ converges to $\bar{f}$, as obviously does the sequence $\{s_n\}$.

The binomial distribution may also be brought under this theorem. Here $P_n(t) = \binom{t}{n} p^n q^{t-n}$, t a positive integer, $p + q = 1$, $0 < p < 1$. Then $\mu = tp$, $\sigma^2 = tpq$. The transformation (7) again leads to the normal distribution (6). The Tauberian theorem is still $\sqrt{n}.a_n = 0(1)$. Euler summability is included in this, being the special case $p = q = \frac{1}{2}$. Note the corollary that *if* 0(1) *is replaced by* $o(1)$ *in the theorem, then the condition on the second moment may be omitted.* If $\sqrt{n}.a_n \to 0$, then the limiting difference quotient exists and is zero, so that every limiting function is a constant without appeal to probability arguments. The summability condition inhibits subsequences tending to different limits. The structure of this and other such probability-Tauberian theorems shows that the difference between the $0(1)$ and $o(1)$ theorems should not be very great, in spite of the formidable difficulties of technique in passing from the original Tauber to the Littlewood theorem.

For $\phi(t)$ an entire function of infinite order, no such completely general result exists. There remains the case of a finite radius of convergence, which may be assumed to be unity without loss of generality. The critical value is $t^* = 1$, approached from the left. The series for $\phi(t)$ must diverge here, and $t = 1$ be a singularity of $\phi(t)$, otherwise $P_n(t)$ would not tend to zero, and the summability method would fail. Now the gap theorems, with the unit circle a natural boundary of $\phi(t)$, show that no unrestricted general result is to be obtained for this case, though extensions such as the "high indices" theorem are possible by specialization. We proceed to consider, in several stages, such generality as is possible.

For the Abel-Tauber theorem, $\phi(t) = 1/(1 - t)$. The probability distribution is the geometric progression, or the "regular absorption" law. It is better to transform the parameter to $z = -1/\log t$, with $z^* = \infty$. Then we have

$$s'(z) = (1 - e^{-1/z}) \Sigma s_n e^{-n/z}. \tag{8}$$

The bracketed expression is asymptotic to $1/z$, so that the transformation is simply $x = n/z$. The same transformation is to be utilized when $\phi(t) = 1/(1 - t)^k$, where k is any positive integer; or, more generally, whenever $b_n \sim n^{k-1}$. The limiting integral is of the form:

$$s = \frac{1}{\Gamma(k)} \int f(x)\, e^{-x}\, x^{k-1}\, dx; \quad k > 1. \tag{9}$$

The formula can be extended to fractional $k \geq 0$, by simple comparison of the orders of magnitude, term by term. For $(C, 1)$ summability, take $P_n(z) = 1/z$ for $n \leq z$; $P_n(t) = 0$, $n > z$. The limit is the uniform distribution: $dF = dx$, $0 \leq x \leq 1$; $dF = 0$ elsewhere. The transformation is again $x = n/z$, though z passes to the critical value infinity only through positive integral values. The (C, k) summability leads to the form

$$k \int_0^1 f(x)\,(1 - x)^{k-1}\, dx. \tag{10}$$

The equivalence between Cesàro and Hölder sums of the same order may be used to deduce the same ultimate limit-form for Hölder means. For Riesz means R_k, the integral in (10) is actually used in direct fashion. In these last cases, there is no infinite power series for some function $\phi(t)$ as in (4), though equivalent summability procedures in series could be constructed. The limiting process is more direct, and can make use of the approximation to (piecewise) monotonic series by means of the corresponding integrals. Thus for $\phi(t) = -\log(1-t)$, $b_n = 1/n$ (which is R. A. Fisher's probability distribution for genes, insect species caught in traps, etc.), the transformation can only be $x = n/z$, but the lower limit becomes 0; the integrand is $f(x)$ $(\exp\text{-}x)/x$ so that if one passes to the limit, the limiting integral is improper. Nevertheless, the Tauberian argument holds, and all the previous results may be obtained by restriction to any x-interval to the right of zero. This is further necessary as $x = 0$ will in general be the solitary point of discontinuity for $f(x)$. Similarly for $b_n = 1/n(\log n)^k$, or $1/n^k$, or $1/n \log n (\log\log n)^k \ldots$ etc. with $k \leq 1$. We have thus proved:

THEOREM 2: *For a $\phi(t)$ in (4) with finite circle of convergence, the Tauberian theorem—if one exists—can only be that $\Sigma\, a_n$ converges to s if it is summable to s, the squared sequence summble to s^2, and $na_n = 0(1)$. The theorem does hold when $\phi(t) = 1/(1-t)^k$, or when b_n tends monotonically to zero in the expansion of ϕ, as well as for all the Cesàro and Hölder means. The condition for Riesz means would be*

$$\frac{\lambda_n a_n}{\lambda_n - \lambda_{n-1}} = 0(1).$$

4. It is seen that the possible types of Tauberian theorems are rather limited. *The condition on the second moment may always be omitted if 0 is replaced by o. For continuous summability, with integration in place of the discrete sum, the place of a_n is taken by the derivative, or the difference quotient, when s_n is replaced by a continuous function.*

It might be thought that the peculiar way in which the set of functions $f(x, t)$ is constructed could lead to convergence without any intervening summability condition. For hypothetical conditions of type $n^k a_n = 0(1)$, $k > 1$, convergence would have been trivially obtained. But this is not true for the weaker conditions which alone are possible for classical Tauberian theorems, as is seen by the harmonic series $a_n = 1/n$, where the sequences are unbounded. For bounded sequences, a counter-example is constructed by taking $a_n = \pm 1/n$, the sign being positive for the first N terms, negative till $n = N^2$, and so on in blocks of $N^r - N^{r-1}$ consecutive terms, with a fairly large N. The sequences then oscillate between zero and log N approximately, without being summable. The situation, and the role of the condition $na_n = 0\,(1)$ may be illustrated for all summability procedures where the larger coefficients are of mutually comparable order, by consideration of $(C, 1)$ summability. Here, we assume without loss of generality that the sum is zero. This may happen by limit-points of opposite signs balancing out, which is prevented by the condition on the second moment. Assume therefore that the $(C, 1)$ summable series has eventually only non-negative sequences. For convergence, $a_n \to 0$ is necessary, which means that for the sub-sequence of " crest " terms $|\alpha_\nu|$ must tend to zero monotonically. The sequence $\{s_n\}$, being bounded, must have at least one cluster point; the summability condition shows that zero must be one of the cluster points. Suppose s_n exceeds some fixed value $h > 0$ infinitely often so that the series does not converge. The least effect on the average in the $(C, 1)$ sum would be when some s_ν is close to zero, and the subsequent terms are all $+ a_\nu$ till h is exceeded, and then $- a_\nu$ till the sequence returns to the neighbourhood of 0 again. The minimum contribution to the Cesâro sum would then be $p\,(p + 1)\, a_\nu/(\nu + 2p)$, where $2p$ is the number of terms in this grouping. However, $p = h/a_\nu$, so that the Cesàro sum has at least the value $h\,(h + a_\nu)/(2h + \nu a_\nu)$. Thus, if na_n exceeds any preassigned value, it would be possible to have an arbitrarily small contribution to the $(C, 1)$ sum from the terms away from 0. With bounded na_n, the only way whereby the (limiting) sum must necessarily vanish is for h to be zero, so that no other limit point than 0 is possible for $\{s_n\}$. The condition $na_n = 0\,(1)$ includes $a_n \to 0$; the case where s_n may have either sign is covered *a fortiori* by the same condition, which suffices for the extreme case. This elementary analysis shows that the structure of the 0 and o results are substantially equivalent for all cases under Theorem 2. It supplies a motivation for the na_n, which appears as the ' best possible ' restriction.

It remains to show for the general case that the condition on the second moment may always be discarded. Lemma 3 and the Tauberian condition yield at least one subsequence from $f(x, t)$ which tends to a continuous function $f(x)$. Suppose that at x_0, there were at least one limiting value other than $f(x_0)$. Then another subsequence can be found tending to that value at x_0. It may be supposed that the new value is greater than $f(x_a)$. From this sequence, another infinite subsequence may be chosen to tend to a limit at a point x_1, and this process may be continued over a denumerable set of points, dense over some suitable interval containing x_0. The Lipschitz condition on $f(x, t)$ then gives a continuous limit function $h(x) > f(x)$ over some interval about $x = x_0$. However, the *sum*, the proper limit of $s'(t)$ in (1), exists regardless of how $t \to t^*$. Also, $P_n(t)$ tends, by hypothesis, to a continuous probability distribution. These data lead to a contradiction, for the range of the limiting integral may be divided into several different intervals; one (which includes x_0) can be made to yield two different contributions to the sum, from $h(x)$ and $f(x)$, while everywhere else, we agree to take $f(x)$ alone. This is permissible inasmuch as our limiting process was by finite intervals; the contribution to the sum could be calculated from any infinite sub-set of functions, separately for each interval. Therefore, one and only one limiting function $f(x)$ can exist.

From the method of construction, we have $\overline{\lim}\, f(x_0, t) = \overline{\lim}\, s_n$, and $\underline{\lim}\, f(x_0, t) = \underline{\lim}\, s_n$, because σx_0 or $z x_0$ must pass through every large integral value. Therefore, if $f(x_0, t) \to f(x_0)$, the sequence $\{s_n\}$ must converge to $f(x_0)$ and, moreover, $f(x) \equiv f(x_0)$. So, we have

THEOREM 3: *The condition on the second moment may be dropped from the general Tauberian theorems 1 and 2, whenever the limiting probability density exists, and is not zero over any sub-interval.*

The main results of this paper were presented in a lecture to the Mathematical Institute, Academia Sinica, Peking, on May 10, 1957.

REFERENCES

1. Loomis, L. H. .. *An Introduction to Abstract Harmonic Analysis*, New York, 1953, 148.

2. Wiener, N. .. *The Fourier Integral* (Cambridge, 1933).
Pitt, H. R. .. *Proc. Lond. Math. Soc.*, 1938, **44**, 243–88 changes Wiener's technique in the direction that probability theory would require.

3. Bochner, S. .. *Harmonic Analysis and the Theory of Probability* (University of California, 1955).

4. Knopp, K. .. *Theory and Application of Infinite Series* (London, 1928), Chap. XIII.
Hardy, G. H. .. *Divergent Series* (Oxford, 1949), particularly Chap. VII and Chap. XII. Our power series used for summability exclude the "high indices" theorem and the like.

5. Kosambi, D. D. .. "Characteristic properties of series distributions," *Proc. Nat. Inst. Sci. India*, 1949, **15**, 109–13. Earliar in *Math. Student*, 1940-**8**, 151–55.

6. Osgood, N. F. .. *Math. Annalen*, 1910, **69**, 71; *cf. Ann. Math.*, 1912-13, **14**, 152–53.