Degree conferred 30 June, 1932.

Ph. D. Thesis. April 1932.

On Relativistic Cosmology, and on the Definition of Distance in General Relativity.

by

Edinburgh.

ON RELATIVISTIC COSMOLOGY, AND ON THE DEFINITION OF DISTANCE IN GENERAL RELATIVITY.

PART I. HISTORICAL INTRODUCTION.

§1. THE COSMOLOGICAL PROBLEM OF RELATIVITY.

A great deal of attention has been directed in recent years to consideration of the metrical character of the Space-Time continuum as a whole, apart from the local distortions which, in accordance with General Relativity, are associated with the presence of matter in condensations such as stars and nebulae. The problem, which we may call the Cosmological problem of Relativity, is to find possible forms for the fundamental tensor g_{ij} specifying the metric $ds^2 = g_{ij}dx^i dx^j$ applicable to Space-Time at large and agreeing with or accounting for the facts of astronomy. In more popular language, if we regard the Space-Time manifold as a four-dimensional hypersurface immersed in flat (pseudo-Euclidean) space of higher dimensions, the problem is to investigate the general shape and dimensions of this surface.

"According to the General Theory of Relativity, the metrical character (curvature) of the four-dimensional

Space-Time continuum is defined at every point by the matter at that point and the state of that matter. Therefore, on account of the lack of uniformity in the distribution of matter, the metrical structure of this continuum must necessarily be extremely complicated. But if we are concerned with the structure only on a large scale, we may represent matter to ourselves as being uniformly distributed over enormous spaces, so that its density of distribution is a variable function which varies extremely slowly. Thus our procedure will somewhat resemble that of the geodesists who, by means of an ellipsoid, approximate to the shape of the earth's surface, which on a small scale is extremely complicated."(1)

The problem has been tackled from two complementary points of view.

(i) Mathematicians have set themselves to investigate the theoretically possible cosmologies and to deduce their physically observable consequences.

(ii) Relevant astronomical data have been collected and compared with the theoretical findings, thus providing

1. A.Einstein: Sitz.Preuss.Akad. Wiss. (1917) 150
translated: "The Principle of Relativity" (Methuen 1921) p.183.

practical tests for the suggested cosmologies, and introducing numerical magnitudes into the discussions.

In the first part of my Thesis I shall review briefly the developments of this subject in the last sixteen years. This is not intended to be a mathematical presentation of the subject, but a historical introduction to the discussions that follow.

§2. EINSTEIN'S CYLINDRICAL WORLD.[1]

In the older views of Space and Time, the Cosmos was regarded as infinite in all directions and uncurved both spatially and temporally; the Space-Time fusion of Special Relativity did not affect this view, but in General Relativity some modification was necessary. The difficulty of assigning boundary conditions[2] at spatial infinity for the fundamental tensor g_{ij}, without either destroying the relativity of inertia or disregarding the observed equilibrium of the stellar system (i.e., the fact that all the stars have not long ago escaped from us to infinity[3]) led Einstein[4] in 1917 to abandon the older view.

He proposed to regard the Universe as "closed"

---

1. Called "Spherical" by some writers.
2. Regarding the g_{ij} as the relativistic gravitational potential, and the field equations of gravitation as analogous to Poisson's equation $\nabla^2\phi = 4\pi\rho$, these are analogous to the boundary conditions which are required in addition to Poisson's equation to make ϕ determinate in the ordinary Newtonian Potential theory.
3. The predominating red-shifts which have since been observed in the spectra of distant nebulae (interpreted as enormous velocities of recession) would seem at first sight to indicate that these objects are in fact escaping to infinity at prodigious speeds. The validity of this deduction from the observations will be discussed later. In 1917 the known facts pointed to the conclusion that the relative motions were on the whole very small and directed at random.
4. loc. cit.

with respect to its spatial dimensions. In such a world, space being finite, the stars would have no chance of escaping to infinity; and, there being no boundary, the question of boundary conditions does not arise.

This necessitated a modification of Einstein's 1916 law of gravitation, his field-equations

$$G_{\mu\nu} = -\kappa\left(T_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} T\right) \qquad (1)$$

being replaced by

$$G_{\mu\nu} - \lambda g_{\mu\nu} = -\kappa\left(T_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} T\right), \qquad (2)$$

where the "cosmical constant" λ must be very small — small enough not to upset the established agreement with observation.[1] (κ is a constant depending on the choice of units.)

With this law, Einstein showed that a spatially closed universe was possible; his solution both preserved the relativity of inertia and permitted stellar equilibrium.

Assuming an ideally uniform distribution of matter, the spatial sections will be of constant curvature, analogous to the circular sections of a cylinder with an

---

1. i.e., the fact that (1) accounts for the observed deviations of planetary motion from Newtonian predictions.

extra curved dimension. That is, any spatial section may be represented as a hypersphere

$$R^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

in a four-dimensional Euclidean space with metric

$$d\sigma^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 .$$

Eliminating the extra variable x_4 and transforming to [illegible] polar coordinates, the spatial metric can be expressed as

$$d\sigma^2 = dr^2 + R^2 \sin^2 \frac{r}{R} \left(d\theta^2 + \sin^2 \theta . d\phi^2 \right) .$$

Thus Einstein was led to suggest the cosmology ("Einstein's cylindrical world") determined by the metric

$$\left. \begin{aligned} ds^2 &= dt^2 - d\sigma^2, \\ \text{where } d\sigma^2 &= dr^2 + R^2 \sin^2 \frac{r}{R} \left(d\theta^2 + \sin^2 \theta . d\phi^2 \right) . \end{aligned} \right\} \quad (3)$$

This is a solution of the equations (2) if:

$$\lambda = \frac{1}{R^2} \quad (4)$$

$T_{44} = \rho_0 g_{44}$ is the only non-zero component of the material Energy-tensor $T_{\mu\nu}$. (5)

where ρ_0 = the proper density of space $= \frac{2}{\kappa R^2}$. (6)

Equation (5) means that the matter in the world is supposed to be at rest; or in other words that the [illegible] pressure, which (disregarding pressure of radiation) is $\frac{2}{3}$ of the total Kinetic Energy of the matter, vanishes;

and the fact that, in the coordinates (3), material particles (following geodesic world-lines) can remain permanently at rest at any point of space, (conforming with the observed fact that the relative velocities of the stars are small compared with that of light) shows that this choice can be regarded as having physical significance, such particles being at rest relative to the stellar system.

Equation (6) provides a guage for the spatial radius. In this connection, Hubble's estimate[1] of the density of space may be quoted. From a statistical study of 400 of the nearer extra-galactic nebulae, which he found to be distributed roughly uniformly in space, he estimated the mean density as 1.5×10^{-31} gm.cm.$^{-3}$, and deduced

$$R = 8.5 \times 10^{28}\,\text{cm} = 2.7 \times 10^{10}\,\text{parsecs} = 8.6 \times 10^{10}\,\text{light-years}$$

as the appropriate radius of Einstein's universe. The corresponding value of the cosmical constant is

$$\lambda = 1.4 \times 10^{-58}\,\text{cm.}$$

---

1. Astrophysical Journal 64 (1926) 368. This estimate gives the density of space in our neighbourhood to a distance of 10^8 light-years, and may be expected to be correct to within a factor 100.

§3. DE SITTER'S SPHERICAL WORLD.

In the same year (1917), de Sitter[1] drew attention to an alternative cosmology (the "de Sitter world") which is also a solution of the modified field equations (2). In this, the time-direction is curved like the spatial directions, imaginary time being homogeneous with the spatial dimensions. The whole continuum is a manifold of constant curvature, analogous to the surface of a sphere with two extra dimensions; but as one of these extra dimensions is imaginary, sections involving real time are hyperbolae instead of circles.

It can be represented as a hyper-pseudo-sphere[2]

$$\mathcal{R}^2 = x_0^2 + x_1^2 + x_2^2 + x_3^2 - x_4^2$$

immersed in a five-dimensional Euclidean space with the metric

$$-ds^2 = dx_0^2 + dx_1^2 + dx_2^2 + dx_3^2 - dx_4^2 .$$

Putting $\frac{x_1}{x_0} = x$, $\frac{x_2}{x_0} = y$, $\frac{x_3}{x_0} = z$, $\frac{x_4}{x_0} = u$,

we get

$$\frac{ds^2}{\mathcal{R}^2} = \frac{du^2 - dx^2 - dy^2 - dz^2}{1 + x^2 + y^2 + z^2 - u^2} + \frac{(u\,du - x\,dx - y\,dy - z\,dz)^2}{(1 + x^2 + y^2 + z^2 - u^2)^2} . \qquad (7)$$

---

1. Proc. Amsterdam. 20 (1917) 230 ; Monthly Notices R.A.S. 78 (1917) 3.
2. The prefix "hyper-" refers to the extra dimensions, "pseudo-" to the hyperbolic changes of sign. The symbol "$\mathcal{R}$" is used to distinguish this 4-dimensional radius from the spatial radius designated "R".

Putting
$$\left.\begin{aligned} x &= \operatorname{sech}\frac{t}{R}\tan\frac{r}{R}\sin\theta\cos\phi \\ y &= \operatorname{sech}\frac{t}{R}\tan\frac{r}{R}\sin\theta\sin\phi \\ z &= \operatorname{sech}\frac{t}{R}\tan\frac{r}{R}\cos\theta \\ u &= \tanh\frac{t}{R} \end{aligned}\right\} \qquad (8)$$

we get :
$$ds^2 = \cos^2\frac{r}{R}\cdot dt^2 - dr^2 - R^2\sin^2\frac{r}{R}\left(d\theta^2 + \sin^2\theta\cdot d\phi^2\right). \qquad (9)$$

This is a solution of the field equations (2) if :
$$\left\{\begin{aligned} \lambda &= \frac{3}{R^2} \qquad (10) \\ T_{\mu\nu} &= 0 \qquad (11) \end{aligned}\right.$$

It will be seen that this metric (9) differs from the metric (3) of Einstein's world only in the coefficient $g_{00} = \cos^2\frac{r}{R}$. de Sitter interpreted this difference as giving rise to a slowing up of time at a distance from the origin. An atomic clock which registers equal intervals ds of proper time will have a period $ds = \cos\frac{r}{R}\cdot dt$ when it is at rest at the point (r, θ, ϕ), as compared with its period ds = dt when it is at rest at the origin. This would appear astronomically as a red-shift of the spectral lines of distant stars or nebulae, and would be interpreted as due to a positive radial velocity.

de Sitter regarded it however as a distance-effect, quite distinct from the Doppler effect due to relative motion between the source of light and the observer, since the atoms in the distant star would actually vibrate more slowly than corresponding atoms in the observer's locality. No such effect being predictable in the Einstein world, he proposed that the existence or non-existence of a predominating red-shift in the spectra of distant stars and nebulae should be regarded as a test for deciding between the two rival cosmologies. At the time, the available data were meagre, but observations quickly confirmed de Sitter's expectations in a remarkable way, and there was soon no doubt of the predominating red-shift among extra-galactic nebulae. The latest results[1] show that there is a linear correlation between red-shift and distance, corresponding to an apparent velocity of recession of 558 km/sec per 10^6 parsecs, with an uncertainty believed not greater than 20%. This is verified to a distance of 3×10^7 parsecs.

de Sitter's conclusion, however, was based on a misinterpretation. The fallacy was pointed out by Lanczos[2], who showed that (in the de Sitter or any other world) there cannot be a red-shift occurring as a distance-effect apart from the ordinary Doppler effect due to the motions of the star and observer. I shall deal with this point in detail later on[3]. It is partly explained by

---

1. Hubble and Humason, Astrophys. Journ. 74 (1931) 43.
2. Zeit. Phys. 17 (1923) 168.
3. Page 31.

pointing out that there is not the same reason for ascribing physical significance to the coordinate system (9) as we found for the system (3) in the Einstein world. Material objects cannot remain _at rest_ in the spatial coordinates of (9) (except at the origin), because by such motion they would not be describing geodesics in space-time. Thus deSitter's calculation of the red-shift is vitiated.

Nevertheless, the correlation of red-shift and distance can be explained on the hypothesis of de Sitter's world and utilised to estimate the appropriate radius. Numerous writers have discussed this and reached discordant conclusions. It seems to me that a great deal that has been written on it is open to criticism. I deal with the question in detail ~~on page~~ in § 17.

§4. COMPARISON OF THE EINSTEIN AND DE SITTER WORLDS.

The Einstein and de Sitter worlds differ in two important respects (among others).

It has already been shown from equation (5) that the matter in the Einstein world must be regarded as being at rest. Consequently Einstein's world is only applicable to the real world in so far as we agree to neglect the individual motions of stars and atoms.

On the other hand, equation (11) shows that de Sitter's world is empty. It must be regarded only as an idealised background for the universe, which would be modified —perhaps in the direction of Einstein's solution— by gravitational distortions superposed on it by the presence of matter. We may of course imagine material particles moving about in the de Sitter world, but they must be supposed devoid of mass. As we have seen, such particles cannot remain at rest in the spatial coordinates of (9), except at the origin.

We may express these points of contrast by saying that

(i) in Einstein's world there is matter but no motion;

(ii) in deSitter's world there is motion but no matter.

The two cosmologies are thus ideal forms, and it is natural to suggest that the real world is somehow intermediate between them. We shall see later that there is reason to suppose that this is so, and in fact that the Universe is expanding from a form resembling Einstein's world in the past to a form resembling de Sitter's world in the future.

§5. FRIEDMAN'S NON-STATIONARY COSMOLOGIES.

The possibility of some long scale process of cosmic evolution suggests that we should look for a non-stationay solution of the field-equations; i.e. for a cosmology which (like Einstein's) involves a "cosmic time" and spherical spatial sections, but whose spatial radius varies slowly with the time. A family of cosmologies of this type was found by Friedman[1] in 1922.

Starting from Einstein's assumptions

(i) the gravitational field-equations (2);

(ii) $T_{44} = \rho_0 g_{44}$ is the only non-zero component of the Energy-tensor $T_{\mu\nu}$;

(iii) the spatial sections of the world are spaces of constant curvature;

he found the stationary solutions of Einstein and deSitter as isolated special cases, and the general non-stationary solution:

$$\left.\begin{aligned} ds^2 &= dt^2 - d\sigma^2 \\ \text{where } d\sigma^2 &= R^2(t).\left(dx_1^2 + \sin^2 x_1 . dx_2^2 + \sin^2 x_1 \sin^2 x_2 \, dx_3^2\right) \end{aligned}\right\}, \quad (12)$$

where the function R(t) is given by the elliptic integral:

---

1. Zeit. Phys. 10 (1922) 385

integral:

$$t - t_0 = \int_{R_0}^{R} \frac{dR}{\sqrt{\frac{1}{3}\lambda R^2 - 1 + \frac{1}{3}\cdot\frac{\alpha}{R}}} \qquad (13)$$

wherein R_0 is the radius at an instant t_0, and α is a constant connected with the mass of the universe. No information about λ (analogous to Einstein's relation[2] $M = \frac{\pi}{2\lambda^2}$) is given by the assumptions; and, depending on the magnitude of λ, Friedman found three possible types of world:

(i) a 'monotonic' world in which the spatial radius increases[1] indefinitely from zero;

(ii) a 'monotonic' world in which the spatial radius increases[1] indefinitely from a finite 'creation' radius;

(iii) a 'periodic' world in which the spatial radius is a periodic function of the time.

---

1. Similar monotonic contracting universes are also possible.

2. Equivalent to (6).

§6. LEMAITRE'S EXPANDING UNIVERSE.

Not very much attention was paid to this paper of Friedman's, and his results were rediscovered independently by Lemaître[1] in 1927. Neglecting the second of the above assumptions, and consequently taking into account the <u>pressure</u> in the universe (p = 2/3 of the Kinetic Energy due to motion, + pressure of radiation), Lemaître arrived at a more general relation correcting <u>R</u> and <u>t</u>, viz.

$$t = \int \frac{dR}{\sqrt{\frac{1}{3}\lambda R^2 - 1 + \frac{1}{3}\frac{\alpha}{R} + \frac{\beta}{R^2}}} \quad (14)$$

where α and β are constants. ($\frac{\alpha}{KR^3}$ is the density, $\frac{\beta}{KR^4}$ is the pressure.)

For comparison with astronomy, however, Lemaitre (in his original paper) decided to neglect the pressure ($\beta = 0$) as being in any case very small; and further, in the absence of other information about the value of Einstein's cosmical constant λ, to assume that $\alpha = \frac{2}{\sqrt{\lambda}}$. The corresponding cosmology ("Lemaître's Expanding Universe") is of the second of Friedman's three types, and the assumption is equivalent to postulating that the 'creation' radius $R_0 = \frac{\alpha}{2}$ shall be given by $\lambda = \frac{1}{R_0^2}$, just as in Einstein's world.

---

1. Annales de la Soc. Sci. de Bruxelles 47A (1927) 49, English translation: Monthly Notices R.A.S. 91 (1931) 483.

(Cf. equation 4.) In other words Lemaître supposes the universe to be expanding from an initial form resembling Einstein's. A consequence of this assumption is that equation 13 (i.e. 14 with $\beta = 0$) can be integrated in the form:

$$\left.\begin{aligned} t &= R_0\sqrt{3}\,\log\frac{1+x}{1-x} + R_0\log\frac{\sqrt{3}\,x-1}{\sqrt{3}\,x+1} + C \\ \text{where } x^2 &= \frac{R}{R+2R_0}\;. \end{aligned}\right\} \qquad (15)$$

On investigating the Doppler effect in this Expanding Universe, taking into account the known correlation (page 10) of red-shift with distance, together with Hubble's estimate (page 7) of the density of space, Lemaître arrived at numerical values for the present radius and the creation radius of the universe:

$$R = 1\cdot 83 \times 10^{28}\text{ cm.} = 6 \times 10^{9}\text{ parsecs} = 2 \times 10^{10}\text{ light-years};$$
$$R_0 = 8\cdot 5 \times 10^{26}\text{ cm.} = 2\cdot 7 \times 10^{8}\text{ parsecs} = 9 \times 10^{8}\text{ light-years}.$$

A considerable amount of further investigation has been done on this model of the Universe by Lemaître, Eddington and others, and these estimates have been modified a little. Eddington[1] drew attention to a conclusion which follows from Lemaître's work, viz. that the Einstein world is unstable, and would have a tendency

---

1. Monthly Notices R.A.S. 90 (1930) 668.

to depart from equilibrium and start either expanding or contracting after Lemaître's model. It must of course be understood that the Einstein world is a complete space-time cosmology, and cannot change *in time* into any other cosmology. In saying that a universe has expanded from the Einstein configuration, we mean that a frustrum of the universe between two spatial sections at some past epoch was indistinguishable from a frustrum of the Einstein world, but that thereafter the spatial sections have been expanding. The fact is that if a spatial universe of constant curvature found itself for a time "in equilibrium" like the Einstein world as regards the constancy of its radius, and if it were disturbed from equilibrium in any way, then it would have a tendency to go on either expanding or contracting, and thus to generate a Lemaître universe. (Or at least one of Friedman's more general cosmologies.) This is what we mean by saying that Einstein's world is unstable, and it is supposed that this may be the actual history of our Universe. Eddington

---

suggested that the departure from equilibrium might be initiated by the formations of condensations of matter, as in nebulae; this possibility has been the subject of investigations [1] by McCrea and McVittie. It is at present too early for me to give any final conclusions on this and the many other points of view from which the Expanding Universe has been considered. Indeed it is perhaps inevitable that a discussion of this problem should always end inconclusively. A quotation giving the conclusions reached in a recent paper [2] by Eddington sums up as well as possible the present state of the problem. Subject to certain provisos, it appears that

"the history of expansion of the Universe resolves itself into a gradual transition from Einstein's to de Sitter's world. . . . the expansion has now proceeded so far that the de Sitter model gives much the better approximation; but this depends entirely on the truth of our estimate that the actual mean density of the Universe is much less than 10^{-28} gm/cm^3. Admitting this,

---

1. Monthly Notices R.A.S. 91 (1930) 128, etc.
2. Ibid. 90 (1930) 676.

no great change is required in current theories which have assumed a de Sitter universe, providing that they do not concern themselves with early history."

§7. ROBERTSON'S COSMOLOGY.

Before closing this introduction, mention should be made of the more general cosmology proposed by Robertson[1] in 1929, and investigated in a series of papers by Tolman[2]. It is equivalent to the metric (12) with R(t) as an arbitrary function, and is thus a very general form of non-stationary cosmology. It arises (see §9 – ~~cosmic coordinates~~, page 27) from general considerations as to the nature of space-time, without reference to any gravitational field-equations. The idea of this generalization is that, since we are more sure of the general homogeneity and isotropy of space than of the exact form of the law of gravitation, it is better to base our cosmological theory on the former only. Of course, there are many investigations (like considering the distortion produced in such a cosmology by the presence of electrons or material condensations) which cannot be performed without assuming some field-equations.

---

1. Proc. Nat. Acad. Sci. 15 (1929) 822
2. Ibid. 16 (1930) 320, 409, 511, 582 Phys. Rev. 38 (1931) 1758, etc.

PART II. THEORETICAL PRINCIPLES.

With this general introduction to relativistic cosmology from a historical point of view, I now proceed to discuss certain logical points which seem to me to be of importance. In doing so, I hope to clear up some misunderstandings which have been responsible for many of the inconsistent conclusions which have been published, especially in connection with the de Sitter world.

§8. COSMIC TIME.

It was originally urged against Einstein's world that by the introduction of 'cosmic time', it restored the absoluteness of Space and Time, which ~~General~~ Special Relativity had so nicely abolished. The same objection applies to the Expanding Universe, in which there is a natural definition of 'at rest'. There is however no real conflict with Relativity. The existence of a unique separation of Space and Time is due to the artificial conditions we have imposed, viz. the complete homogeneity of the pressure and density. Any lack of homogeneity in the Universe would give rise to an uncertainty in this space-time separation, somewhat analogous to the uncertainties of the Quantum Theory.

In the real Universe, there is reason to believe that relative stellar velocities are all small compared with that of light. That is, their world-lines form a bundle of fibres which are all more or less parallel. The individual fibres may vary a little (or even a lot) but even if(with the expansion of the Universe) they diverge considerably, there is still not much doubt of the direction of the whole bundle. This direction we take as the direction of cosmic time.

Thus, even if the real Universe is not perfectly homogeneous as regards pressure and density, nevertheless the approximate stillness of the stellar system defines, within certain small limits of uncertainty, a unique space-time separation. A perfect cosmological theory would take this uncertainty into account; and it is recognised that in trying to fit an ideal Expanding Universe to the real Universe, we are neglecting just this uncertainty in the direction of cosmic time.

§9. "COSMIC" COORDINATES.

It should be remembered that from a mathematical point of view, material particles are to be regarded as lines, not as points. The same applies to observers. They are in fact geodesics in space-time. Now, in the real world, we know that there are local gravitational fields superposed on the cosmical background. In the neighbourhood of such a distortion a particle is deflected as if by a field of force. The stars and nebulae are too far apart to feel each other's individual gravitational fields, but an observer on the earth is carried in ellipses round the sun, so that his world-line is nothing like a geodesic in the Expanding Universe of Lemaitre. Nevertheless he is careful to correct his astronomical observations by taking this motion into account, thus replacing himself by an ideal observer whose motion is a geodesic in the cosmical background. He goes further and makes corrections for the solar motion, thus adopting a particular cosmic time (determined by the motion of our own Galaxy) quite different from his own personal direction in Space-Time. Future observers may even learn to make corrections for the peculiar motion of our own Galaxy among the nebulae; but in view of the inevitable uncertainty which exists in any choice of cosmic time, it is doubtful if this would make much difference.

If we are going to regard the observer and the nebulae as being spatially at rest, we must take care to use coordinates in which this is possible. Let us find the general condition that a coordinate system may satisfy this requirement.

Consider the general metric

$$ds^2 = g_{ij}dx^i dx^j , \qquad (i,j = 0,1,2,3)$$

where x^0 is the time coordinate and x^1, x^2, x^3, are spatial coordinates. We wish to find the condition that the line

$$x^1 = \text{const.}, \; x^2 = \text{const.}, \; x^3 = \text{const.},$$

may be a geodesic orthogonal[1] to the spatial section x^0 = const..

To be a geodesic, the line must satisfy

---

1. This assumption of orthogonality, leading to $g_{01} = g_{02} = g_{03} = 0$, is equivalent to postulating symmetry in past & future time, or to postulating that the velocity of light shall be the same in opposite directions.

$$\frac{d^2x^i}{ds^2} + \left\{ {i \atop j\,k} \right\} \frac{dx^j}{dx} \frac{dx^k}{ds} = 0 .$$

Hence $\left\{ {\alpha \atop 0\,0} \right\} = 0$ for $\alpha = 1, 2, 3$.

Hence $[0\,0, \alpha] = 0$; i.e. $\frac{\partial g_{0\alpha}}{\partial x^0} - \frac{1}{2} \frac{\partial g_{00}}{\partial x^\alpha} = 0$. (16)

Further, the direction $(dx^0, 0, 0, 0)$ of the line must be orthogonal to every spatial direction $(0, dx^1, dx^2, dx^3)$.

Hence $g_{0\alpha} = 0$. (17)

Comparing (16) and (17), we see that the required conditions are that $g_{0\alpha} = 0$ and that g_{00} should be a function of x^0 only. By a change of the time-coordinate, viz. putting $dt = \sqrt{g_{00}}\,dx^0$, we can arrange for g_{00} to be unity. This simply means that we are making our coordinate time coincide with the proper time of particles which are spatially at rest, so that there is a natural reason for identifying it with cosmic time. The metric now becomes $ds^2 = dt^2 - d\sigma^2$, (18)

where $d\sigma^2$ is a spatial metric[1]; and we see that this is the necessary form of the metric in any space-time separation defining a cosmic time, subject to the correctness of our original assumption that material

---

1. Non-static: i.e. its coefficients can be functions of the time.

objects can remain spatially at rest.

Such coordinates, effecting a natural separation of space and time and defining universal simultaneity, might be called 'cosmic coordinates'.[1]

To complete the logical development, it may be observed that:

(i) the further assumption of an ideally uniform distribution of matter means that space is homogeneous and isotropic, so that $d\sigma^2$ is the metric of a space of constant curvature; this leads to the general cosmology (12), equivalent to that proposed by Robertson;

(ii) the further assumption that space-time satisfies the gravitational equations (2) leads to Friedman's cosmologies.

(iii) the further assumption that space is expanding from an initial Einstein configuration leads to Lemaître's Expanding Universe;

(iv) the further assumption that the spatial radius is constant leads to Einstein's Cylindrical World.

---

1. cf. Hilbert, Math. Ann. 92 (1924) 15;
Robertson, Proc. Nat. Acad. Sci. 15 (1929) 822.
Tolman, Ibid. 16 (1930) 320.

§10. THE CURVATURE OF SPACE.

The assumption of the constant curvature of space is interpreted differently by Eddington.[1] He regards it as a necessity philosophically inherent in the nature of things, independent of the distribution of matter. For the radius of space provides in any locality a natural unit of length; in the absence of any other standard, he supposes that the ultimate constituents of matter have to adjust themselves automatically so that their dimensions, in terms of this natural standard, are everywhere the same; our own units of length (cms., parsecs, etc.) naturally follow the electrons in this adjustment; so that when we come to measure the radius of space in terms of these units, of course we find it is constant.

I confess I do not see the force of this argument. It seems to me that if electrons are going to respond to space-curvature at all, they would find it much easier to appreciate the strong curvatures of local

---

1. Mathematical Theory of Relativity, C.U.P. 1930, §66.

gravitational fields, which mask the barely noticeable curvature of the cosmological background. So that it seems more likely that the dimensions of electrons are prescribed in terms of some cosmical standard length such as $\frac{1}{\sqrt{\lambda}}$; this would be related to the mean curvature of space; but the actual curvature would be greatest in those regions of space in which most matter happens to be congregated, and constant only in so far as matter may be regarded as uniformly distributed.

§11. THE DE SITTER WORLD IN "COSMIC" COORDINATES.

It will be observed that all the cosmologies which we have ~~so far~~ been considering have been of the general form (18), with the exception of the de Sitter world (9). By a transformation of coordinates, however, the de Sitter world also can be put in this form. Thus, if in place of (8), we put

$$\left.\begin{aligned} x &= \tan \chi_1 \sin \chi_2 \cos \chi_3 \\ y &= \tan \chi_1 \sin \chi_2 \sin \chi_3 \\ z &= \tan \chi_1 \cos \chi_2 \\ u &= \tanh \frac{t}{R} \sec \chi_1 \end{aligned}\right\} \qquad (19)$$

we obtain from (7) Lanczos's[1] form of the de Sitter world:

$$\left.\begin{aligned} ds^2 &= dt^2 - d\sigma^2, \\ \text{where} \quad d\sigma^2 &= R^2 \cosh^2 \frac{t}{R} \left(d\chi_1^2 + \sin^2\chi_1\, d\chi_2^2 + \sin^2\chi_1 \sin^2\chi_2\, d\chi_3^2 \right). \end{aligned}\right\} \qquad (20)$$

This shows that the de Sitter world can be regarded as one of Friedman's cosmologies, arising when the density is put equal to zero.. For if in (13) we put $\alpha = 0$ and integrate, we obtain

$$R = \sqrt{\frac{3}{\lambda}} \cosh \sqrt{\frac{\lambda}{3}}\, t \;;$$

thus, since by (10) R has the value $\sqrt{\frac{3}{\lambda}}$, (12) reduces to (20).

[see note (2), page 8]

---

1. Zeit. Phys. 17 (1923) 168.

20 being of the general form 18 , material particles can remain at rest at any point of space[1]; and we can regard the coordinates as having physical significance, this space-time separation corresponding to that determined by the approximate stillness of the stellar system. Other coordinates can of course be used for purposes of calculation; but for purposes of physical interpretation which depend on the distinction between space and time, we must evidently rely on a "cosmic" representation such as 20 . The weird phenomena which de Sitter found in his world from a study of the metric 9 to be associated with the "mass horizon" of singularities $\frac{\Lambda}{R} = \frac{\pi}{2}$, where time stands still, do not arise at all in these coordinates. Further, there is no red-shift occurring as a distance effect; this is true of any metric of the form 18 . For, since $g_{00} = 1$ we have ds = dt for any particle at rest; thus atomic clocks recording equal intervals ds will also record

---

1. This only means that their coordinates remain fixed. Their mutual distances will increase with the spatial expansion of the universe.

equal times dt when at rest in any part of the universe; in other words, similar atoms will vibrate everywhere with the same frequency. Consequently, any observed red-shift in a stellar spectrum must be interpreted as a Doppler effect due to the motions of star and observer.

At the same time care is needed in interpreting Doppler effects, in interpreting for instance a positive Doppler effect as being necessarily proportional to a positive radial velocity, and a predominance of positive Doppler effects as ~~being~~ necessarily indicating expansion of the universe. The fact is that there is ambiguity in the usage of such words as distance and velocity, and it is necessary to see clearly in what sense we are using them before we can safely interpret the Doppler effects. In view of the importance of such interpretation when fitting cosmological theories to astronomical data, I propose to discuss the matter at considerable length.

§12. DEFINITIONS OF DISTANCE.

"Distance" in ordinary geometry, analogous to "interval" in space-time, is an invariant relating to two points. I have previously emphasized the fact that in Relativity observers and particles are not points, but lines. "Distance" in Relativity relates primarily to two lines; but since it is understood to refer to a particular instant and may vary with the time, this invariant will have relevance to a particular point of time, and so to a particular point on one of the world-lines, say the observer's. If we are going to regard distance as an invariant relating to two world-points, it will be necessary to specify a definite point on the other world-line—we will call it the "star"'s—to go with the first. It is in this correlation of points on the two world-lines that the ambiguity arises. There are two possibilities for the correlation, and several possible definitions of distance.

(i) We may correlate those points on the two world-lines which lie on the same spatial section t = const., and define the distance σ as the geodesic distance $\int d\sigma$ or $i\int ds$ integrated between the two world-points. This may be appropriately called "spatial distance", being simply the ordinary (Riemannian) distance in the space whose metric is $d\sigma^2$. The definition assumes the existence of cosmic time (and accordingly a metric of the form (18)). It depends only on the two world-points: i.e. it is not concerned with the motion of the star and the observer as indicated by the inclination of their world-lines to the time axis. And it is symmetrical: it does not matter which is the star and which is the observer.

Corresponding spatial radial velocity may be defined by the rate of increase of spatial distance with respect to cosmic time, and appears as a divergence of the two world-lines. On account of the fact that the spatial metric is in general non-static, so that the spatial distance involves the time, this divergence can appear between two particles which are both spatially at rest. A general divergence of all world-lines representing particles at rest is implied by expansion of the universe. Between a particular pair of particles not at rest, the divergence due to expansion may be masked if their peculiar motions are big enough, but the divergence would appear statistically if the peculiar motions were not too big and were directed at random.

In the universes of de Sitter, Lemaitre and Friedman, spatial distance ~~distance~~ is given (from (12)) by

$$\sigma = R(t) . \int \sqrt{dx_1^2 + \sin^2 x_1 . dx_2^2 + \sin^2 x_1 . \sin^2 x_2 . dx_3^2}$$

integrated between the two world-points. Thus the residual divergence is $\dfrac{d\sigma}{dt} = \dfrac{R'}{R}\,\sigma$ (21)

which (for a fixed t) increases proportionally to the spatial distance between the two points, and at sufficiently great distances may be expected to be comparable with the divergence due to the peculiar motions.

If we were to assume that σ and $\frac{d\sigma}{dt}$ coincide to a first approximation with astronomical distance and Doppler velocity, we could identify this formula with the correlation quoted on page 10 ; this would give $\frac{R'}{R} = 6.2 \times 10^{-28} cm^{-1}$. But we have to investigate the corrections to be applied to the astronomical measurements before accepting this result.

(ii) Alternatively, we might correlate those points on the two world-lines which lie on the same null-geodesic. It is of course to such correlation that astronomical measurements of distance refer; for astronomical observations are necessarily made by rays of light, i.e. along null-geodesics in space-time. The geodesic distance between the two points is then zero, but there are other ways in which the distance can be defined. Unlike spatial distance, it will in any case be unsymmetrical; for the null-geodesic must be such as to represent light going from the star to the observer and not vice versa.

(a) We could define distance as the lapse of cosmic time between the two world-points, i.e. the time taken by light from the star to reach the observer.

(This of course is to be multiplied by c, the velocity of light, if in the units used c is not unity.) This may be appropriately called "temporal distance". Like spatial distance, it assumes the existence of cosmic time and does not depend on the motions of the star and the observer.

(b) Although it would seem that temporal distance is the thing that is really contemplated by astronomers, that they aim at measuring, it does not follow that it is the quantity that they actually measure. With the nearer stars the corrections to be applied would probably be negligible in any case, but we are mainly concerned with the very distant extragalactic nebulae. Since it is these "astronomical distances" which are correlated with Dop Doppler effects, it is important (for cosmological investigations) that we should know how they are to be represented mathematically. The question has been considered by several writers already, but not so far as I am aware, completely satisfactorily, except with reference to one or two particular metrics. I propose to discuss it in considerable detail, and to do so primarily with reference to a perfectly general Riemannian space-time. I have two reasons for considering a general space-time rather than (say) Lemaitre's cosmology. Firstly, it is not yet established definitely that the universe is of Lemaitre's form; secondly, it may be desirable in future investigations to take into account the effect of the gravitational fields of the star and the observer on the estimation of distance, and a general metric provides for this.

§13. THE DEFINITION OF DISTANCE IN ACCORDANCE WITH ASTRONOMICAL METHODS.

The methods by which astronomers compute the distances of the extragalactic nebulae[1] depend ultimately on a comparison of the absolute and apparent brightness of some object, on the assumption that brightness decreases with the square of the distance. The object observed may be a nebula itself, or some involved star of recognisable type. If the star is a Cepheid variable, the absolute brightness can be inferred with considerable accuracy from the period of its variations. Less accurate methods of estimating absolute brightness involve conclusions from spectral type, or the rough assumptions (supported by such evidence as is available) that similar nebulae, or the brightest stars to be found in similar nebulae, are of about the same order of absolute brightness.

In a recent paper[2], Whittaker suggested that this principle— the comparison of absolute and apparent

---

1. v.Hubble, Astrophys.Journ. 64 (1926) 321.
2. Proc. R.S., A 133 (1931) 93.

brightness on the assumption that brightness decreases with the square of the distance—translated into the language of differential geometry, should be adopted as a definiton of distance in general Riemannian space-time; and applied this definition to the de Sitter world. Previously Tolman[1] investigated this "distance" for ~~the~~ a line-element equivalent to (12). It would seem, however, that neither Whittaker nor Tolman correctly translated the astronomical procedure.

In accordance with this definition, the distance between the world-point A (the star at the instant when light was emitted) and the world-point B (the observer at the instant when the light is received) will only be defined when the interval AB is zero, i.e. when A and B are points on the same null-geodesic, so that a ray of light can go from A to B, and, further, when a definite world-direction, specifying the observer's motion, is associated witht the point B,

---

1. Proc.Nat.Acad.Sci.16 (1930) 511.

for observers moving differently would in general make different observations of apparent brightness.

Whittaker expresses the definition as follows: "We consdier a thin pencil of null-geodesics (rays of light) which issue from A and pass near B . This pencil intersects the observer B's 'instantaneous 3-dimensional space', giving a 2-dimensional cross-section; the 'distance AB' is then defined to be proportional to the square root of this cross-section." The factor of proportionality is to be independent of the actual position of B on the null-geodesic, and is to be such that when A and B are close together, "AB" shall reduce to distance as measured in B's local "space".(1)

Now there is a certain ambiguity involved in this factor of proportionality. We recall that "absolute brightness" is defined astronomically as the brightness of the star as it would appear to an observer at some standard distance from it. The standard distance is usually taken as 10 parsecs, the nebular distances to

---

1. The "instantaneous 3-dimensional space" or "local space" of an observer is that portion of the 3-dimensional space orthogonal to the ~~proper time~~ direction of his world-line which is in his immediate neighbourhood.

be measured being millions of parsecs. Thus, when we talk of the absolute brightness of a star, we are really postulating a fictitious observer at a world-point A_1 on the same null-geodesic AB (so that he observes the same pencil of light), but in the neighbourhood of the star; and we are imagining the estimate of brightness to be made by him. If Θ and θ are the areas of the cross-sections of the same thin pencil as observed at B and A_1, and if δ is the distance AA_1 as measured in the local space of the fictitious observer, then the "distance" Δ is given by $\Delta^2 = \frac{\Theta}{\theta}\delta$. Thus the factor of proportionality is $\frac{\delta}{\sqrt{\theta}}$. This depends not only on the position of A and the direction AA_1, but also on the unspecified world-direction associated with A_1; i.e. on the motion of the fictitious observer. How are we to consider him to be moving?

We might take the world-line of our fictitious observer to be "parallel" to the world-line of the original observer, in the sense of interescting it in an arbitrarily chosen "point at infinity" K. The two world-lines will then coincide when B coincides with A_1, giving $\Theta = \theta$; so that Δ reduces, as it should, to δ .

This is in fact what Whittaker does. In applying his definition to the de Sitter world, he uses the coordinates (7), in which the geodesics have linear equations; the observer's world-line is taken "parallel to the axis of u "; and the factor of proportionality is evaluated by considering the star and the observer to be close together, *the observer's world-line being still parallel to the axis of u* .

~~The restricted invariance of this parallelism is emphasized by Whittaker. Thus, his final formula expresses the distance in terms of two quantities , ; and he makes it clear that these are invariants only for such transformations as preserve parallelism to the axis of u -- i.e. for such transformations as preserve the same point at infinity K on the observer's world-e line. It should however be pointed out that K is an~~

His final formula for distance is as follows: The distance between a star whose world-coordinates are (uxyz) and an observer whose world-line is the geodesic $x=x_0$, $y=y_0$, $z=z_0$, where x_0, y_0, z_0, are constants, is

$$\Delta = R\frac{\sin\rho}{\cos(\sigma+\rho)} \qquad (22)$$

where

$$\cos\rho = \frac{1+xx_0+yy_0+zz_0}{(1+x^2+y^2+z^2)^{\frac{1}{2}}(1+x_0^2+y_0^2+z_0^2)^{\frac{1}{2}}}, \quad (0\leq\rho\leq\pi)$$

and

$$\sin\sigma = \frac{u}{(1+x^2+y^2+z^2)^{1/2}}, \quad (-\tfrac{\pi}{2}\leq\rho\leq\pi)$$

the radicals being taken positively.

The quantities ρ and σ are invariants only for such transformations as preserve parallelism with the axis of u —i.e. for such transformations as preserve the same point at infinity K on the observer's world-line. It should however be pointed out that K is an arbitrary point on the observer's world-line. For, by a suitable transformation of the type:

$$u' = \frac{u \cosh\gamma - \sinh\gamma}{u \sinh\gamma + \cosh\gamma} \quad ; \quad x', y', z' = \frac{x,\ y,\ z}{u \sinh\gamma + \cosh\gamma},$$

(which does not affect the form of the metric), any point $u = -\coth\gamma$ on a given geodesic can be transformed into the point at infinity.

It is clear that when we specify the world-line of the fictitious observer in this way, our definition of distance will depend not only on the positions and motions of the star and the observer, but also on the quite arbitrarily chosen point at infinity K on the observer's world-line. In fact, Whittaker's "invariant" σ is actually arc tan $(\cos \frac{AK}{R})$, involving the arbitrary geodesic distance AK.

For an effective definition of distance, we should have to make some special choice of K, on which the otherwise arbitrary "absolute brightness" depends.

~~A natural suggestion suggestion would be~~

No such choice however will yield a definition of distance which accurately translates the astronomical procedure. For we recall that the factor of proportionality $\frac{\delta}{\sqrt{\theta}}$, which is the brightness at unit distance, is intended to be an intrinsic property of the star. If we take the world-line of the fictitious observer to depend in any way on the position or motion of the original observer, then it is clear that we are not really dealing in absolute brightness at all. The only alternative is to suppose that the fictitious observer is moving with the star. This will be justified if astronomical "absolute brightness" can be identified with the "proper brightness" of the star—that is to say, if there are no corrections (depending on the observer) which ought to be applied to the astronomers' estimate of absolute brightness, when we take Relativity considerations into account.

Now, in the case of a Cepheid variable, as already pointed out, the "absolute brightness" is calculated by a well-known relation from the observed period of variation. On account of relative motion between the star and the observer, this observed period is not the "proper" period of the star's variations; but the appropriate Doppler correction is revealed by spectroscopic analysis. In practice the correction is not applied; this is not from neglect of Relativity considerations, but simply because (in view of the wide margin of error present in any case) it is too small to affect the final result. If we suppose this correction to be applied, and neglect errors of ~~obser~~ observation, then there is no doubt that the astronomical absolute brightness is the proper brightness of the star. The same conclusion would apply to other methods of estimating absolute brightness (in which however the margin of error is much larger).

Therefore we must give our fictitious observer the motion of the star.

It is important to remark that when the star and the (original) observer are close together, the "distance" as we have now defined it does **not** reduce to distance as measured in the observer's local "space". I shall in fact show that it reduces to distance as measured in the local space of the star, and further that in space-time of constant curvature (the de Sitter world), even when the star is remote from the observer, the __observer__'s motion is not involved. But in the most general case, our definition of distance involves the motions or world-directions of both star and observer.

Tolman (loc. cit.) follows a different procedure from ours. Instead of comparing the apparent brightness of the star with a similar observation made by a fictitious __observer__ in the neighbourhood of the __star__, he compares it with a similar observation made on a fictitious __star__ in the neighbourhood of the __observer__. Here again, my criticism is that the comparison brightness is not an intrinsic property of the star; for it depends on the shape of space-time in the neighbourhood of the __observer__ at the instant when he receives the light. Thus, in Tolman's calculations, it involves the function $g(t)$

of the cosmic time which figures in his metric evaluated at the instant of reception of light by the observer.

In the same paper, Tolman considers an alternative definition of distance, based on a comparison of the absolute and apparent dimensions of a distant star or nebula. In this case we have to consider rays of light issuing from different points of the star A, and arriving simultaneouoly at the observer B—i.e. a thin pencil of null-geodesics passing near A and converging at B. Here again, Tolman employs for comparison observations made on a fictitious star in the neighbourhood of the observer. But let us, as before, employ for comparison observations made on the star itself by an observer moving with it, so that "absolute size" will be an intrinsic property of the star. If the apparent dimensions are measured by the solid angle subtended by the star at B, i.e. by the area intersected by the pencil on a small sphere round B of radius δ, and if this is compared with the actual area under observation, as measured by someone moving with the star—then our original definition is exactly reversed, and the formula for this distance Δ' will be given by interchanging the roles of star and observer in the formula we obtain for Δ.

I now proceed to apply these considerations to a general Riemannian space-time, obtaining a general formula for "distance" as measured astronomically.

RT III. FORMULAE AND APPLICATIONS.

§14. GENERAL FORMULA FOR ASTRONOMICAL "DISTANCE" IN RIEMANNIAN SPACE-TIME.

Take the world-point A (the star at the instant when a light-pulse is emitted) for origin, and choose a system of coordinates such that

(i) the null-geodesics of the pencil we are considering (passing through A and close to B) have <u>linear</u> equations;

(ii) if the metric (for time-like intervals) is

$$ds^2 = g_{ij}dx^i dx^j, \qquad (i,j=0,1,2,3) \qquad (23)$$

then $\frac{\partial g_{ij}}{\partial x^k} = 0$ in the neighbourhood of the origin A. (24)

Such a choice of coordinates is always possible; we can for instance take "geodetic" coordinates[1]; the conditions (i) and (ii) will then be satisfied, and we shall have further :-

(iii) <u>all</u> geodesics through the origin have linear equations;

(iv) $\left\{ {i \atop jk} \right\} x^j x^k = 0$ at every point.

But the conditions (i) and (ii) will be sufficient for our purposes.

---

1. These are the "Riemannian co-ordinates" described in ch. II of Eisenhart, Riemannian Geometry (Princeton 1926).

Let the null-geodesic AB be: $x^i = \xi^i \lambda$, (25)

where ξ^i are four constants satisfying[1]: $g_{ij}\xi^i\xi^j = 0$; (26)

and let the "fictitious observer" A_1 be at the point of parameter λ. Let v be the unit vector in the direction of his world-line, which we are taking to be the world-direction of the star, so that his local space is:

$$v_i dx^i = 0 \ . \qquad (27)$$

That is, a displacement dx^i, to be recognised by him as spatial, must satisfy (27).

We wish to find the linear element, and hence the element of area, in the wave-front of the light from the star as observed at A_1 — i.e. the two-dimensional cross-section of the space (27) by the cone of null-geodesics from A. That is, we wish to find the distance between $A_1(x^i)$ and A_1' $(x^i + dx^i)$, when A_1', besides being in the hyperplane (27), is taken to be on a null-geodesic adjacent to (25).

For the ray AA_1' (which is a null-geddesic adjacent to AB), we have, by differentiating (25) and (26) and remembering (24):

1. Throughout this section, unless otherwise indicated, g_{ij} denotes the fundamental tensor at the point A. Its value at B will be denoted $(g_{ij})_B$.

$$\begin{cases} dx^i = \xi^i \, d\lambda + \lambda \, d\xi^i , & (28) \\ \xi_i \, d\xi^i = 0 & (29) \end{cases}$$

where $\xi_i = g_{ij}\xi^j$.

From (27) and (28), putting $v_i \xi^i = K$:

$$K \, d\lambda + \lambda \, v_i \, d\xi^i = 0 . \qquad (30)$$

This gives $d\lambda$. Substituting in (28):

$$dx^i = \lambda \left(\delta^i_j - \frac{v_j \xi^i}{K} d\xi^j \right) \qquad (31)$$

where $\delta^i_j = 0$ or 1 according as i, j are unequal or equal.

Substituting in (23): $$-A_1 A_1'^2 = \lambda^2 g_{ij} \left(\delta^i_l - \frac{v_l \xi^i}{k} \right) d\xi^l . \left(\delta^j_m - \frac{v_m \xi^j}{k} \right) d\xi^m \qquad (32)$$

$$= \lambda^2 \left(g_{lm} - \frac{v_l \xi_m + v_m \xi_l}{k} \right) d\xi^l d\xi^m , \qquad (33)$$

the remaining terms $\frac{\lambda^2 g_{ij} v_l v_m \xi^i \xi^j}{k} d\xi^l d\xi^m$ vanishing on account of (26).

So far we have only specified the ratios of the constants ξ^i. To fix their absolute values, let $a_i \xi^i = 1$, so that

$$a_i \, d\xi^i = 0 , \qquad (34)$$

where a_i are four constants.

Take coordinates η^2, η^3 in the wave-front, such that

$$\eta^\alpha = a^\alpha_i \xi^i \qquad (\alpha = 2, 3) .$$

Ultimately we shall take $\eta^2 = \xi^2$, $\eta^3 = \xi^3$; but for the present a^α_i are any eight constants.

Then:

$$d\eta^\alpha = a^\alpha_i \, d\xi^i \qquad (\alpha = 2, 3) \qquad (35)$$

Solving the four equations (29), (34), (35) for $d\xi^i$, we get:

$$d\xi^i = A^i_\alpha \, d\eta^\alpha , \qquad (36)$$

where: $A^i_\alpha = \dfrac{\text{minor of } a^\alpha_i \text{ in } \Xi}{\Xi}$, Ξ being the determinant

$$\begin{vmatrix} \xi_0 & \xi_1 & \xi_2 & \xi_3 \\ a_0 & a_1 & a_2 & a_3 \\ a^2_0 & a^2_1 & a^2_2 & a^2_3 \\ a^3_0 & a^3_1 & a^3_2 & a^3_3 \end{vmatrix}$$

Substituting (36) in (33), we get for the linear element in the wave-front: $-A_1 A_1'^2 = \gamma_{\alpha\beta}\, d\eta^\alpha d\eta^\beta$, (37)

where: $$\gamma_{\alpha\beta} = \lambda^2 \left(g_{lm} - \frac{v_l \xi_m + v_m \xi_l}{k}\right) A^l_\alpha A^m_\beta . \qquad (38)$$

Now, from the definition of A^i_α, $\xi_l A^l_\alpha = 0$, (39)

this being equal to a determinant with two rows the same. Thus, all the terms in (38) which involve v vanish, and

$$\gamma_{\alpha\beta} = \lambda^2 g_{ij} A^i_\alpha A^j_\beta . \qquad (40)$$

So the element of area, $\theta \equiv (\gamma_{22}\gamma_{33} - \gamma_{23}^2)^{\frac{1}{2}} d\eta^2 d\eta^3$, (41)

is independent of the motion v of the fictitious observer.

Of course, it does not follow that Θ, referring to the original observer at B, is independent of his motion motion: for when we deal with an observer remote from the star, the ξ_l in (33) and (38) $= (g_{lm})_B \xi^m$, while the ξ_i in (29) and the determinant Ξ $= (g_{ij})_A \xi^j$; so that (39) no longer holds. Consider however the case when the star at A and the observer at B are close together, so that we can take the fictitious observer also to be at B. Then Θ reduces to θ, being now independent of the observer's motion; and the distance $\Delta \equiv \sqrt{\frac{\Theta}{\theta}} . \delta$ reduces to δ, which is distance in the local space of an observer moving with the Star, as asserted above (page 48).

From (40) and (41) we have:

$$\theta = \lambda^2 \left[g_{ij} A_2^i A_2^j . g_{kl} A_3^k A_3^l - g_{ij} A_2^i A_3^j . g_{kl} A_2^k A_3^l \right]^{\frac{1}{2}} d\eta^2 d\eta^3$$

$$= \lambda^2 \phi^{\frac{1}{2}} d\eta^2 d\eta^3 \qquad (42)$$

where $$\phi = g_{ij}\, g_{kl}\, A_2^i A_3^l \left(A_2^j A_3^k - A_2^k A_3^j \right). \qquad (43)$$

Remembering that a_i^α are arbitrary constants, let us now take: $a_i^2 = 0,0,1,0 \,; \quad a_i^3 = 0,0,0,1 \,;$

thus: $\eta^2 = \xi^2 , \; \eta^3 = \xi^3 , \; d\eta^\alpha = d\xi^\alpha .$

Then:

$$\Xi = \begin{vmatrix} \xi_0 & \xi_1 & \xi_2 & \xi_3 \\ a_0 & a_1 & a_2 & a_3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = a_1 \xi_0 - a_0 \xi_1 .$$

Hence, writing $a_i \xi_j - a_j \xi_i \equiv A_{ij}$, we have: (44)

$$\left. \begin{aligned} A_2^0 &= \frac{A_{21}}{A_{10}} , \; A_2^1 = \frac{A_{02}}{A_{10}} , \; A_2^2 = 1 , \; A_2^3 = 0 , \\ A_3^0 &= \frac{A_{31}}{A_{10}} , \; A_3^1 = \frac{A_{03}}{A_{01}} , \; A_3^2 = 0 , \; A_3^3 = 1 . \end{aligned} \right\} \qquad (45)$$

(43) can now be written: $$\phi = \frac{g_{ij}.g_{kl}.A^{ij.kl}}{2A_{10}^2} , \qquad (46)$$

where $A^{ij.kl}$ = the second minor of A_{ij}, A_{kl} in the skew-symmetric determinant

$$\begin{vmatrix} A_{00} & A_{01} & A_{02} & A_{03} \\ A_{10} & A_{11} & A_{12} & A_{13} \\ A_{20} & A_{21} & A_{22} & A_{23} \\ A_{30} & A_{31} & A_{32} & A_{33} \end{vmatrix} .$$

I content myself with verifying only one pair of terms in (46), showing that they agree with (43), and asserting that the other terms may be similarly verified.

Consider the terms $i, j, k, l, = 0, 0, 1, 1 ; 1, 1, 0, 0$.

In (43), these

$$= g_{00}\, g_{11} \left[A^0_2 A^1_3 \left(A^0_2 A^1_3 - A^1_2 A^0_3 \right) + A^1_2 A^0_3 \left(A^1_2 A^0_3 - A^0_2 A^1_3 \right) \right]$$

$$= \frac{g_{00}\, g_{11}}{A_{10}{}^4} \left(A_{21} A_{03} - A_{02} A_{31} \right)^2$$

$$= \frac{g_{00}\, g_{11}}{A_{10}{}^4} \left[(a_2 \xi_1 - a_1 \xi_2)(a_0 \xi_3 - a_3 \xi_0) - (a_0 \xi_2 - a_2 \xi_0)(a_3 \xi_1 - a_1 \xi_3) \right]^2$$

$$= \frac{g_{00}\, g_{11}}{A_{10}{}^4} \left(a_2 a_0 \xi_1 \xi_3 + a_1 a_3 \xi_0 \xi_2 - a_0 a_3 \xi_1 \xi_2 - a_1 a_2 \xi_0 \xi_3 \right)^2$$

$$= \frac{g_{00}\, g_{11}}{A_{10}{}^4} \left(a_2 \xi_3 - a_3 \xi_2 \right)^2 \left(a_0 \xi_1 - a_1 \xi_0 \right)^2$$

$$= \frac{g_{00}\, g_{11}\, A_{23}{}^2}{A_{10}{}^2}$$

$$= g_{00}\, g_{11} \frac{A^{00.11} + A^{11.00}}{2\, A_{10}{}^2},$$ which are the corresponding terms in (46).

Other pairs of terms may be similarly but more easily compared, thus verifying (46).

From (42) and (46), $\theta = \lambda^2 \Psi_A^{\frac{1}{2}} \dfrac{d\eta^2 d\eta^3}{A_{10}}$, (47)

where $\Psi_A \equiv \frac{1}{2} g_{ij} g_{kl} A^{ij.kl}$.

$\delta \equiv$ distance of star as measured in the local space of the fictitious observer

= the projection of AA_1 on ~~the~~ his time-axis

(for, since A and A_1 are on the same null-geodesic, their space-separation = their time-separation)

$= v_i x^i = \lambda v_i \xi^i$. (48)

To find a formula for Δ, it only remains to find Θ. Remembering that the four quantities a_i (see (34)) which occur in the definition (44) of A_{ij}, and hence in (47), are at our disposal, let us take them to be the covariant components of the unit-vector in the direction of the observer's world-line. We proceed to apply to Θ the method used to evaluate θ. Using (34), we have now:

in place of (30) : $d\lambda = 0$;

in place of (31) : $dx^i = \Lambda d\xi^i$, where Λ is the parameter of B on (25);

in place of (32) : $-BB'^2 = \Lambda^2 (g_{ij})_B \, d\xi^i d\xi^j$;

in place of (37) & (40): $-BB'^2 = \Gamma_{\alpha\beta} \, d\eta^\alpha d\eta^\beta$, where $\Gamma_{\alpha\beta} = \Lambda^2 (g_{ij})_B A^i_\alpha A^j_\beta$;

in place of (47) : $\Theta = \Lambda^2 \Psi_B^{\frac{1}{2}} \dfrac{d\eta^2 d\eta^3}{A_{10}}$ (49)

where $\Psi_B = \frac{1}{2} (g_{ij})_B (g_{kl})_B A^{ij.kl}$.

The "distance AB" is given by $\Delta = \left(\frac{\Theta}{\theta}\right)^{\frac{1}{2}}\delta$. Hence, from (47), (48), + (49) we have:

$$\Delta = \Lambda v_i \xi^i \left(\frac{\psi_A}{\psi_B}\right)^{1/4}$$

Now $\Lambda \xi^i = X^i$, the coordinates of B;

and hence $\Lambda \xi_i = X_i$, where $X_i = g_{ij}X^j$.

We shall evidently get the same result if we replace $A^{ij.kl}$ by quantities proportional to them, so we may aswell say that

$$A_{ij} \equiv a_i X_j - a_j X_i \qquad (50)$$

in place of (44).

We can now express as follows our final formula for distance estimated by a comparison of absolute and apparent brightness:-

We are considering two world-points A,B on a null-geodesic representing a ray of light from A to B. Choose coordinates as on page 51, with origin A; and let v, a, be the vectors specifying the motions of the star and the observer (i.e. unit vectors directed along their respective world-lines). Form the quantities $X_i = (g_{ij})_A X^j$. From these, form the skew-symmetric determinant of the quantities $A_{ij} = a_i X_j - a_j X_i$, and let $A^{ij.kl}$ denote the second minor of A_{ij}, A_{kl}. Evaluate the expression $\psi = \frac{1}{2} g_{ij} g_{kl} A^{ij.kl}$ at both the points A,B.

Then the "distance AB" is: $$\Delta = v^i X_i \left(\frac{\psi_B}{\psi_A}\right)^{1/4}. \qquad (51)$$

It will be shown in the next section that $v^i X_i = \left(1 + \frac{\delta\lambda}{\lambda}\right) . a^i X_i$, where $\frac{\delta\lambda}{\lambda}$ is the fractional red-shift found by the observer in the star's spectrum.

Thus (51) can be written: $$\Delta = \left(1 + \frac{\delta\lambda}{\lambda}\right) a^i X_i \left(\frac{\psi_B}{\psi_A}\right)^{1/4}, \qquad (52)$$
in which form the motion of the star is not involved, except in so far as it produces the Doppler effect.

$\psi = \frac{1}{2} g_{ij} g_{kl} A^{ij.kl}$ can be written in the equivalent form:
$\psi = \frac{1}{2} g^{ij.kl} A_{ij} A_{kl}$, where $g^{ij.kl}$ = the second minor of g_{ij}, g_{kl} in the determinant $|g_{ij}|$. The full expression for ψ is given on the next page.

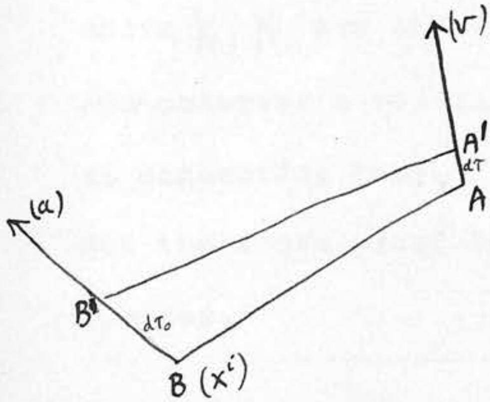
$$
\begin{aligned}
\Psi = {} & A_{23}^2\left(g_{00}\, g_{11} - g_{01}^2\right) + A_{31}^2\left(g_{00} g_{22} - g_{02}^2\right) + A_{12}^2\left(g_{00} g_{33} - g_{03}^2\right) \\
& + A_{01}^2\left(g_{22}\, g_{33} - g_{23}^2\right) + A_{02}^2\left(g_{11}\, g_{33} - g_{13}^2\right) + A_{03}^2\left(g_{11} g_{22} - g_{12}^2\right) \\
& + 2A_{12}A_{31}\left(g_{00}g_{23} - g_{02}g_{03}\right) + 2A_{23}A_{12}\left(g_{00}\, g_{31} - g_{01}\, g_{03}\right) + 2A_{31}A_{23}\left(g_{00}\, g_{12} - g_{01}\, g_{02}\right) \\
& + 2A_{02}A_{03}\left(g_{12}\, g_{31} - g_{11}\, g_{23}\right) + 2A_{03}A_{01}\left(g_{23}\, g_{12} - g_{22}\, g_{31}\right) + 2A_{01}A_{02}\left(g_{31}\, g_{23} - g_{33}\, g_{12}\right) \\
& + 2A_{23}\, A_{01}\left(g_{31} g_{02} - g_{12}\, g_{03}\right) + 2A_{23}A_{02}\left(g_{11} g_{03} - g_{31}\, g_{01}\right) + 2A_{23}A_{03}\left(g_{12} g_{01} - g_{11}\, g_{02}\right) \\
& + 2A_{31}\, A_{01}\left(g_{23} g_{02} - g_{22} g_{03}\right) + 2A_{31}\, A_{02}\left(g_{12} g_{03} - g_{23}\, g_{01}\right) + 2A_{31}\, A_{03}\left(g_{22} g_{01} - g_{12}\, g_{02}\right) \\
& + 2A_{12}\, A_{01}\left(g_{33} g_{02} - g_{23} g_{03}\right) + 2A_{12}\, A_{02}\left(g_{31}\, g_{03} - g_{33}\, g_{01}\right) + 2A_{12}\, A_{03}\left(g_{23} g_{01} - g_{31} g_{02}\right).
\end{aligned}
\tag{53}
$$

We have thus arrived at two equivalent formulae (51), (52) for distance as measured astronomically by comparison of absolute and apparent brightness. With reference to (52), it may be pointed out that neglect of the factor $\left(1+\frac{\delta\lambda}{\lambda}\right)$ would make no appreciable difference to most estimates of nebular distances. In the most distant ~~nebulae~~ nebulae, however, $\frac{\delta\lambda}{\lambda}$ has been observed of the order of magnitude of about $\frac{1}{20}$; and since it is reasonable to suppose that much greater Doppler shifts will be found when more distant nebulae are examined, it is clear that this factor might have to be taken into account in the near future.

The disadvantage of these formulae is that they are not expressed in tensor form, so that they are not at once applicable to any given metric. I hope to remedy ~~this~~ this in a future investigation, and to apply the formulae to Lemaitre's Expanding Universe. Meanwhile, since as we have seen there is reason to believe that the Universe is very nearly of de Sitter's form, it will be interesting to see how these formulae work out when applied to that world. This I shall discuss in § 16.

§15. GENERAL FORMULA FOR THE DOPPLER EFEECT IN RIEMANNIAN SPACE-TIME.

With the coordinates and notation of the last section,



consider a light pulse emitted from A during an interval $d\tau = AA'$ of the star's proper-time; and suppose that its arrival at B occupies an interval $d\tau_0 = BB'$ of the observer's proper-time.

Then the Doppler shift as observed at B is given by:

$$\frac{\lambda + \delta\lambda}{\lambda} = \frac{d\tau_0}{d\tau}.$$

A' is the point $x^i = v^i d\tau$, B' is the point $x^i = X^i + a^i d\tau_0$.

These points are on a null-geodesic whose equation is approximately linear, so that we have approximately:

$$g_{ij}\left(X^i + a^i d\tau_0 - v^i d\tau\right)\left(X^j + a^j d\tau_0 - v^j d\tau\right) = 0 ;$$

or, to the first order, $X_i\left(a^i d\tau_0 - v^i d\tau\right) = 0,$

where as before $X_i = g_{ij}X^j$ (the g_{ij} being evaluated at A).

Therefore : $v^i X_i = \frac{d\tau_0}{d\tau} \cdot a^i X_i$

Therefore :
$$\left(1 + \frac{\delta\lambda}{\lambda}\right) = \frac{v^i X_i}{a^j X_j}, \qquad (54)$$

as was assumed on page 59.

This formula for the Doppler effect in general space-time is equivalent to Lanczos's formula[1]

$$1 + \frac{\delta\lambda}{\lambda} = \frac{\cos\gamma_1}{\cos\gamma_0},$$

where γ_0, γ_1 are the (imaginary) angles which the star's and observer's world-lines make with the null-geodesic AB connecting them. For $\frac{v^i x_i}{a^i x_i} = \frac{v^i \xi_i}{a^i \xi_i} = \frac{\cos\gamma_1}{\cos\gamma_0}$. But the above proof is much shorter than that given by Lanczos.

---

1. Lanczos, Zeit.Phys. 17 (1923) 177, equation (22).

§16. APPLICATION TO THE DE SITTER WORLD.

We take the metric of the de Sitter world in the form (cf. (7))

$$\frac{ds^2}{R^2} = \frac{du^2 - dx^2 - dy^2 - dz^2}{1 + x^2 + y^2 + z^2 - u^2} + \frac{(udu - xdx - ydy - zdz)^2}{(1 + x^2 + y^2 + z^2 - u^2)^2}. \quad (55)$$

The origin of these coordinates is arbitrary, so we can take it to be the point A, i.e. the star at the instant when the light pulse is emitted; it may be shown that all geodesics have linear equations; and the functions $\frac{\partial g_{ij}}{\partial x^k}$ vanish at the origin. Thus the two conditions (i) and (ii) on page 51 are satisfied, and we can proceed to apply the formulae for astronomical distance Δ.

Before doing this I shall find the formula in these coordinates for an invariant which is of importance in the de Sitter world, viz. the minimum geodesic distance of a given world-point from a given world-line. It will be recalled that in Whittaker's discussion of the de Sitter world, two invariants ρ, σ were introduced. The extra invariant arises when a particular point on the world-line is specialised as the point at infinity; but otherwise there is only <u>one</u> invariant of this configuration.

Consider the point (u_1, x_1, y_1, z_1), and the world-line

$$x = au + a', \quad y = bu + b', \quad z = cu + c'. \tag{56}$$

The metric (55) can be integrated, so that if s is the (space-like) interval between $(u_1 x_1 y_1 z_1)$ and $(uxyz)$, i.e. $i\int ds$ integrated along the geodesic joining them, then:

$$\cos\frac{s}{R} = \frac{1 + xx_1 + yy_1 + zz_1 - uu_1}{(1+x^2+y^2+z^2-u^2)^{\frac{1}{2}}(1+x_1^2+y_1^2+z_1^2-u_1^2)^{\frac{1}{2}}}. \tag{57}$$

If $(uxyz)$ is a point on (56),

$$\cos\frac{s}{R} = \frac{Eu + F}{(Au^2+2Bu+C)^{\frac{1}{2}}.G^{\frac{1}{2}}} \tag{58}$$

where

$$\left.\begin{array}{ll} A = a^2+b^2+c^2-1, & E = ax_1 + by_1 + cz_1 - u_1 \\ B = aa'+bb'+cc', & F = a'x_1 + b'y_1 + c'z_1 + 1 \\ C = a'^2+b'^2+c'^2+1, & G = 1 + x_1^2 + y_1^2 + z_1^2 - u_1^2 \end{array}\right\} \tag{59}$$

Differentiating (58),

$$-\frac{1}{R}\sin\frac{s}{R}\cdot\frac{ds}{du} = \frac{(BE-AF)u + CE - BF}{(Au^2+2Bu+C)^{3/2}\,G^{\frac{1}{2}}}$$

Thus s is a minimum when $u = \dfrac{BF-CE}{BE-AF}$

Denoting the minimum value by D, (58) gives:

$$\cos^2\frac{D}{R} = \frac{-AF^2+2BEF-CE^2}{(B^2-AC)\,G}. \tag{60}$$

This is the formula for the invariant D in these coordinates. It will be useful to obtain the corresponding formulae in other systems of coordinates as well.

The metric (9): $ds^2 = \cos^2\frac{r}{R}\,dt^2 - dr^2 - R^2\sin^2\frac{r}{R}\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)$ (61)

is derived from (55) by putting

$$\left.\begin{array}{l} x = \operatorname{sech}\frac{t}{R}\tan\frac{r}{R}\sin\theta\cos\phi \\ y = \operatorname{sech}\frac{t}{R}\tan\frac{r}{R}\sin\theta\sin\phi \\ z = \operatorname{sech}\frac{t}{R}\tan\frac{r}{R}\cos\theta \\ u = \tanh\frac{t}{R} \end{array}\right\} \tag{8}$$

Consider the point (r,θ,ϕ,t) and the line r=0, i.e. x=y=z=0, which is the world-line of a particle at rest at the spatial origin. The equations (8) give the coordinates of the point in our original system and we find from (59) and (60) :-

A=-1, B=0, C=1, E=$-\tanh\frac{t}{R}$, F=1, G= $\operatorname{sech}^2\frac{t}{R}\sec^2\frac{r}{R}$, $\cos^2\frac{D}{R}=\cos^2\frac{r}{R}$.

Hence $D = r$.

Again, the metric (20) : $ds^2 = dt^2 - R^2\cosh^2\frac{t}{R}\left(dx_1^2+\sin^2x_1\,dx_2^2+\sin^2x_1\sin^2x_2\,dx_3^2\right)$ (62)

is derived from (55) by putting

$$\left.\begin{aligned} x &= \tan x_1 \ \sin x_2 \ \cos x_3 \\ y &= \tan x_1 \ \sin x_2 \ \sin x_3 \\ z &= \tan x_1 \ \cos x_2 \\ u &= \tanh\frac{t}{R}\ \sec x_1 \end{aligned}\right\} \quad (19)$$

Consider the point (x_1, x_2, x_3, t) and the line $x_1=0$, which again is x=y=z=0, and is the world-line of a particle at rest at the spatial origin. We find:-

A=-1, B=0, C=1, $E=-\tanh\frac{t}{R}\sec x_1$, ~~E=-tanh t/R~~, F=1, G=$\sec^2x_1.\operatorname{sech}^2\frac{t}{R}$.

Hence: $\cos^2\frac{D}{R} = \dfrac{1-\tanh^2\frac{t}{R}\sec^2x_1}{\sec^2x_1\operatorname{sech}^2\frac{t}{R}} = 1-\cosh^2\frac{t}{R}\sin^2x_1$

i.e. $\sin\frac{D}{R} = \cosh\frac{t}{R}\sin x_1$

In what follows, I shall use D to denote the minimum geodesic distance of the star's world-point from the world-line of the observer; and D_0 to denote the minimum geodesic distance of the observer's world-point from the world-line of the star.

I now proceed to apply the formulae (51), (52) to the metric (55). We have taken the origin at A, so that at A:

$$g_{00} = R^2, \quad g_{11} = g_{22} = g_{33} = -R^2, \qquad (x^0, x^1, x^2, x^3 \equiv u, x, y, z)$$

other components of $g_{ij} = 0$.

Without loss of generality we can take the axis of u for the world-line of the star, so that: $v^i = \left(\frac{1}{R}, 0, 0, 0\right)$; and further we can take B to be the point: $X^i = (u_0, x_0, 0, 0)$.

Since B is on a null-geodesic through the origin,

$g_{ij} x_0^i x_0^j = 0$; i.e. $u_0^2 - x_0^2 = 0$.

Therefore $x_0 = \pm u_0$.

Taking $x_0 = u_0$, we have:

$$\left.\begin{aligned} X^i &= (u_0, u_0, 0, 0), \\ X_i &= (R^2 u_0, -R^2 u_0, 0, 0). \end{aligned}\right\} \qquad (63)$$

Also: $$v^i X_i = R u_0. \qquad (64)$$

$$A_{ij} \equiv a_i X_j - a_j X_i = \begin{pmatrix} 0, & -(a_0 + a_1) R^2 u_0, & -a_2 R^2 u_0, & -a_3 R^2 u_0 \\ (a_1 + a_0) R^2 u_0, & 0, & a_2 R^2 u_0, & a_3 R^2 u_0 \\ a_2 R^2 u_0, & -a_2 R^2 u_0, & 0, & 0 \\ a_3 R^2 u_0, & -a_3 R^2 u_0, & 0, & 0 \end{pmatrix}$$

From (53):

$$\begin{aligned} \Psi_A &= \left(A_{01}^2 + A_{02}^2 + A_{03}^2 - A_{23}^2 - A_{31}^2 - A_{12}^2\right) R^8 u_0^2 \\ &= \left[(a_0 + a_1)^2 + a_2^2 + a_3^2 - a_3^2 - a_2^2\right] R^8 u_0^2 \\ &= (a_0 + a_1)^2 R^8 u_0^2 . \end{aligned} \qquad (65)$$

At B: $$\left.\begin{aligned} &g_{00} = R^2(1 + u_0^2), \quad g_{11} = R^2(-1 + u_0^2), \quad g_{22} = g_{33} = -R^2, \quad g_{01} = -R^2 u_0^2 \\ &\text{other components of } g_{ij} = 0 . \end{aligned}\right\} \qquad (66)$$

$$\begin{aligned} \therefore \quad \Psi_B &= A_{31}^2 g_{00} g_{22} + A_{12}^2 g_{00} g_{33} + A_{01}^2 g_{22} g_{33} + A_{02}^2 g_{11} g_{33} + A_{03}^2 g_{11} g_{22} \\ &\quad + 2 A_{31} A_{03} g_{22} g_{01} - 2 A_{12} A_{02} g_{33} g_{01} \\ &= \left[(a_3^2 + a_2^2)(-1 - u_0^2) + (a_0 + a_1)^2 + (a_2^2 + a_3^2)(1 - u_0^2) + 2(a_3^2 + a_2^2) u_0^2\right] R^8 u_0^2 \\ &= (a_0 + a_1)^2 R^8 u_0^2 . \end{aligned} \qquad (67)$$

From (51), (64), (65), (67): $\Delta = Ru_0$ (68)

It will be observed that $\psi_A = \psi_B$, so that Θ like θ is (69)
independent of any world-direction (cf. page 48) .
Consequently in (68) the world-direction a of the observer is not involved. We can therefore express Δ in terms of the invariant D_0 of the world-point of the observer and the world-line of the star; the resulting formula, being a relation between invariants, will be true generally and applicable to any system of coordinates.

Putting $u_1=x_1=u_0$, $y_1=z_1=a=b=c=a'=b'=c'=0$ in (59) and (60), we find: $A=-1$, $B=0$, $C=1$, $E=-u_0$, $F=1$, $G=1$; $\cos^2\frac{D_0}{R} = 1-u_0^2$.
Hence, comparing (68), $\Delta = R\sin\frac{D_0}{R}$. (70)

On the other hand, introducing the Doppler effect and applying (52), we can express Δ in terms of the invariant D of the world-point of the star and the world-line of the observer. As before, we take A at the origin, and B at $(u_0, u_0, 0, 0)$; but now we take the world-line of the observer at B parallel to the axis of u, that of the star being unspecified. Equations (63), (66), (69) remain true, and we have to evaluate a^1X_1. The observer's world-line is: $x=u_0, y=z=0$. Thus $a^1=a^2=a^3=0$, while a^0 is given by the condition that a be a unit vector, namely $(g_{00})_B(a^0)^2=1$; this gives $a^0 = \frac{1}{R(1+u_0^2)^{\frac{1}{2}}}$.

Hence $a^i X_i = a^0 X_0 = \frac{Ru_0}{(1+u_0^2)^{1/2}}$.

Therefore, by (52) and (69), $\Delta = \left(1+\frac{\delta\lambda}{\lambda}\right)\frac{Ru_0}{(1+u_0^2)^{1/2}}$

Now for the point (0,0,0,0) and the world-line $x=u_0, y=z=0$, (60) reduces to: $\cos^2\frac{D}{R} = \frac{1}{1+u_0^2}$.

Thus:

$$\Delta = \left(1+\frac{\delta\lambda}{\lambda}\right) R \sin\frac{D}{R} \,. \qquad (71)$$

From (70) + (71), $\frac{\delta\lambda}{\lambda} = \frac{\sin D_0/R}{\sin D/R} - 1$. (72)

We have thus two equivalent general formulae for the distance Δ of a star from an observer in the de Sitter world, when estimated by brightness.

(i) (70) gives Δ in terms of the world-point of the observer and the world-line of the star; while

(ii) (71) gives Δ in terms of the world-point of the star, the world-line of the observer, and the Doppler shift in the star's spectrum.

In the coordinates (uxyz) for which the metric is (55):-

(i) the distance of a star describing the world-line $x=au+a'$, $y=bu+b'$, $z=cu+c'$ from an observer at $(u_1 x_1 y_1 z_1)$ is given by (59), (60) + (70);

(ii) the distance of a star at $(u_1 x_1 y_1 z_1)$ from an observer describing the world-line $x=au+a'$, $y=bu+b'$, $z=cu+c'$, is given by (59) (60) & (71).

In the polar coordinates (r,θ,ϕ,t) for which the metric is (61),

(i) if the star is at rest at the spatial origin, and the observer's first coordinate is r_0, then $\Delta = R\sin\frac{r}{R}$; (73)

(ii) if the observer is at rest at the spatial origin, and the star's first coordinate is r, then: $\Delta = \left(1+\frac{\delta\lambda}{\lambda}\right)R\sin\frac{r}{R}$ (74)

In the cosmic coordinates $(x_1 x_2 x_3 t)$ for which the metric is (62),

(i) if the star is at rest at the spatial origin $x_1=0$, and the observer is at (x_1, x_2, x_3, t), then: $\Delta = R\cosh\frac{t}{R}\sin x_1$, (75)

(ii) if the observer is at rest at the spatial origin, and the star is at (x_1, x_2, x_3, t), then: $\Delta = \left(1+\frac{\delta\lambda}{\lambda}\right)R\cosh\frac{t}{R}\sin x_1$. (76)

In the above formulae it is of course understood that the observer and the star are on the same null-geodesic representing light from the star to the observer; thus, when the actual position of one is defined, the position of the other on its world-line is fixed.

Thus, in the cosmic coordinates $(x_1 x_2 x_3 t)$ in the second of the two cases just considered, if the observer's time-coordinate is t_0, then the two points $(000t_0)$, $(x_1 x_2 x_3 t)$ lie on a null-geodesic. The condition for this can be found by transforming to the coordinates (uxyz) and expressing the fact that the interval as given by (57) is zero. The condition is:

$$1 - \tanh\frac{t_0}{R}\tanh\frac{t}{R}\sec x_1 = \operatorname{sech}\frac{t_0}{R}\operatorname{sech}\frac{t}{R}\sec x_1 .$$

Solving for t, we find:

$$\cosh\frac{t}{R} = \cosh\frac{t_0}{R}\cdot\frac{\cos x_1 \pm \sinh\frac{t_0}{R}\sin x_1}{1-\cosh^2\frac{t_0}{R}\sin x_1} .$$

We must take the __earlier__ of the two values of t given by this, the later one corresponding to the instant when the light reflected from the observer reaches the star again. (76) now becomes:

$$\Delta = \left(1+\frac{\delta\lambda}{\lambda}\right) R\cosh\frac{t_0}{R}\cdot\frac{\cos x_1 \pm \sinh\frac{t_0}{R}\sin x_1}{1-\cosh^2\frac{t_0}{R}\sin x_1}\sin x_1 .$$

This gives Δ in terms of the distance coordinate x_1 of the star and the time-coordinate t_0 of the observer at rest at the spatial origin, when he observes the star.

It will be observed that x_1 cannot be taken as an approximation to Δ. On the other hand, as is seen from (73) and (74), in the polar coordinates $(r\theta\phi t)$ the distance coordinate r is a good approximation to Δ, whether we take the observer or the star to be at rest at the origin, provided that $\frac{r}{R}$ and $\frac{\delta\lambda}{\lambda}$ are small.

In accordance with the considerations on pages 49 & 50, the other distance Δ' defined by comparison of absolute and apparent extension, is given from (70) by

$$\Delta' = R \sin \frac{D}{R} \qquad (77)$$

We see that Δ' is independent of the motion of the star. Comparing (71), we find:

$$\Delta = \Delta'\left(1 + \frac{\delta\lambda}{\lambda}\right). \qquad (78)$$

Now for spherical objects of the same absolute size and brightness, placed at different distances from the observer, the apparent brightness varies as $\frac{1}{\Delta^2}$, and the apparent area as $\frac{1}{\Delta'^2}$. Thus by (78) the ratio of the apparent brightness to the apparent area is inversely proportional to $\left(1+\frac{\delta\lambda}{\lambda}\right)^2$; or, the ratio of the square root of the apparent brightness to the apparent <u>diameter</u> is inversely proportional to $1+\frac{\delta\lambda}{\lambda}$. This result was found by Tolman[1] for any non-stationary cosmology of Robertson's general form. (My criticism of his procedure in defining distance does not affect the validity of this result

---

1. Proc. Nat. Acad. Sci. 16 (1930) 519, cf. 593.

of his.) He suggested that a correlation of this form between apparent brightness, diameter and Doppler effect should be looked for in the extragalactic nebulae, as a negative result would indicate that the universe is not even of Robertson's form ~~at all~~. Care is needed in establishing such a correlation in view of the effects of orientation on apparent size; for most of the nebulae appear as ellipses, and probably most of those which appear spherical are really ellipsoids viewed sideways.

I think however that the suggested test might be usefully used to test a hypothesis which is suggested by Hubble's researches.[1] THe correlation which has actually been found between apparent brightness anD size (after taking into account the statistical effects of random orientation) indicates that the square root of the apparent brightness is nearly proportional to the minor diameter; there is reason to believe that the nebulae differ very little in absolute brightness; it has therefoe been inferred that their (absolute) minor diameters differ very little. But their major diameters differ a

---

1. Astrophys. Journ. 64 (1926) 321 .

great deal. This and other considerations suggest that the whole collection of elliptical nebulae arranged in order of increasing ellipticity represents the various configurations of ~~an~~ one originally globular mass expanding equatorially, viewed at random distances and with random orientations, the absolute brightness and minor diameters remaining constant during the expansion. The correlation however is not very good; if the hypotheses are correct, it ought to be improved by taking into account the Doppler effect in the way suggested above. In this case, instead of dealing with all the nebulae at once in seeking the correlation, they would be divided into a number of groups according to gradations of $\frac{\delta\lambda}{\lambda}$; in each group a linear correlation would be sought between [the square root of the absolute brightness] and [the minor diameter ~~multiplied~~ divided by $1+\frac{\delta\lambda}{\lambda}$]; the factor of proportionality should turn out to be the same in all the groups.

§17. THE SIZE OF THE UNIVERSE.

Having arrived at accurate formulae (70) and (72) for distance and the Doppler effect in the deSitter world, agreeing with astronomical procedure, we are in a position to investigate the correlation which is to be expected between the distance and the Doppler effect, and to make comparison with the astronomically determined correlation already quoted (page 10), for the purpose of estimating the radius R. I do not suggest that my investigation supersedes those already made, for the problem has been extensively studied; and until more accurate observational data are established, my "distance" corrections will remain mere theoretical niceties. However, the use of the coordinates (uxyz) makes my approach to the subject simpler than those of Tolman and Silberstein, whose somewhat cumbrous calculations are necessitated by the use of polar coordinates.

In this investigation, it is of course necessary to make some assumption about the distribution and motions of the nebulae. On the observational side, it is known that the nebulae appear uniformly distributed throughout space as far as we can see. It will be as well for us to see what assumptions, on the theoretical side, have been made by the various writers on the subject.

Silberstein's[1] procedure may be described as follows:

He uses the polar coordinates (61), taking the observer to be at rest at the origin; and considers a star describing an arbitrary geodesic. He identifies the coordinate r with the distance. (We have seen from (74) that r is a good approximation to Δ, since $\frac{r}{R}$ and $\frac{\delta\lambda}{\lambda}$ are small.) He finds an equation[2] relating the radial velocity of a particle to its distance in terms of two "star-constants" v_0, r_0. These are respectively the velocity and distance of the star when it is at "perihelion" i.e. at the nearest point to the observer in its whole history) These star-constants are for individual stars quite inexplorable; but for statistical purposes Silberstein assumes that: (i) v_0 is distributed at random among the stars (i.e. its distribution follows the normal error law; (ii) all values of $\frac{r_0}{r}$ between 0 and 1 are equally probable regardless of the value of r itself. The value obtained for R is 3.6×10^7 parsecs ($= 1.2 \times 10^8$ light-years).[3]

---

1. The Size of the Universe. (O.U.P. 1930).
2. Ibid. (90) page 136.
3. The calculation was made before the data quoted on page 10 were available, & was based on other results.

I have two criticisms to make of these assumptions.

(i) The definiton of v_0—the velocity with which the star passes through perihelion—depends on the space-time separation determined by the metric (61). As we have seen, the cosmic separation (62) is the one which has physical significance, and the velocity will be differently defined in these coordinates. I venture to guess that the velocity in the coordinates (61) will always be greater than in the cosmic coordinates. In any case the discrepancy increases consistently with increasing distance from the origin. From the considerations of §9, it is natural to suppose that v_0 in the cosmic coordinates is subject to random distriubtion, in which case it will <u>not</u> be subject to random distribution in the polar coordinates, but will be correlated with the distance.

(ii) Silberstein apparently[1] intends his second assumption to express the fact that (for stars at a given distance r) the perihelia are uniformly distributed over the space within a sphere of radius r round the origin. If so, he is mistaken; for with such a distribution it is the <u>large</u> values of r_0 (and hence of $\frac{r_0}{r}$) which would predominate.

---

1. v. Ibid. page 138.

The uniform distribution of r_0 for a given r is only possible if a preponderance of perihelia are clustered near the origin, i.e. if a majority of the stars are a long way past perihelion, i.e. if there is an average motion of recession. In the absence of any physical explanation to account for the particular distribution of perihelia necessary to distribute $\frac{r_0}{R}$ uniformly, the assumption seems to me dangerously like an ad hoc and not very plausible assumption.

Tolman's investigation[1] is based on the metric

$$ds^2 = \left(1 - \frac{r^2}{R^2}\right) dt^2 - \left[\frac{dr^2}{1 - \frac{r^2}{R^2}} + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2\right].$$

This becomes (61) on substituting $R \sin \frac{r}{R}$ for r. On investigating measurement of distances from an observer at the origin, Tolman decides that the coordinate r agrees exactly with astronomical distance as measured by comparison of absolute and apparent extension; and agrees with that measured by comparison of brightness subject to corrections "not at present within the range of observation"[2]. This is borne out by our equations (71) and (77), where Tolman's r (our $R \sin \frac{D}{R}$) agrees with Δ' and differs from Δ by the factor $\left(1 + \frac{\delta\lambda}{\lambda}\right)$.

Tolman considers four alternative hypotheses about the nebular distributions. The first, "the hypothesis of continual entry", is as follows:

---

1. Astrophys. Journal. 69 (1929) 245.
2. At the time, the largest Doppler effects which had been found were of the order $\frac{\delta\lambda}{\lambda} = \frac{1}{100}$. They have now risen to about $\frac{1}{20}$.

"The uniform concentration of nebulae within our range of observation is produced and maintained by the continuous arrival of new nebulae, approaching perihelion, to make good the deficiency caused by those which have passed perihelion and are leaving the range of observation never again to return."

Tolman finds that this hypothesis (though he regards it as inherently the most probable) is not sufficient to account for the observed correlation of apparent radial velocity and distance, without some ~~further~~ restrictions on the motions of the nebulae; but it leads to the information that R cannot be less than 2×10^8 light-years ($= 6.5 \times 10^7$ parsecs).(1)

Tolman's alternative hypotheses are all rather similar, being three different ways of accounting for the view that the motions of the nebulae are approximately radial; that is, that their world-lines all intersected, or came very close to, the world-line of our own galaxy in the remote past, so that their perihelia are all zero or very small. This leads to the required correlation with $R = 2 \times 10^9$ light-years ($= 6.5 \times 10^8$ parsecs).(1)

1. The calculations were based on a preliminary estimate of the Doppler-distance correlation, before the data on page 10 were accurately established.

however, for reasons which I do not clearly understand, Tolman dismisses these three hopeful hypotheses as being "inherently improbable" or "involving *ad hoc* assumptions", and decides finally that

"the de Sitter line-element for the universe does not appear to afford a simple and unmistakably evident explanation of our present knowledge of the distribution, distances, and Doppler effects for the extragalactic nebulae," although "by the introduction of *ad hoc* assumptions" we could get any desired relation between Doppler effect and distance.

But the suggestion of a common origin of the nebulae and our own galaxy in the remote past is a very attractive hypothesis in my opinion. It accounts simultaneously for the great distances of the nebulae, their consistent recession from us, and their approximate stillness (i.e. the fact that their velocities of recession are all small compared with that of light). For if the nebulae parted company with our galaxy at a remote epoch, their distances must be expected to be immense by now; they would all appear to be receding; and only those which have been moving approximately "parallel" to our own will now be visible.

The hypothesis was studied by Castelnuovo[1], who drew attention to the much greater simplicity of calculation afforded by the coordinates (55). He expressed his conclusion as follows:

"If in the de Sitter world of radius $R = 2 \times 10^9$ light-years, the Milky Way and the extragalactic nebulae were scattered at an exceedingly remote epoch (theoretically $t = -\infty$) by an immense central nebula, thus acquiring a free motion through space, we should receive from their movements precisely the impression which is revealed by spectrum analysis."

The numerical result agrees with Tolman's, both having approximated. Logically, however, Castelnuovo's work is spoilt by the fact that his "distance" does not agree with astronomical distance; ~~in the first stage 55-56~~ it is what I have called "spatial distance" (page 34), being measured in a spatial section of the universe and not along a null-geodesic.

I shall therefore examine the question afresh.

---

1. Monthly Notices R.A.S. 91 (1931) 829.

using the astronomical distance Δ.

In the coordinates (uxyz), for which the metric is (55) we can without loss of generality take the observer's world-line to be $x=y=z=0$, and suppose him to be at the point (0000). From (55) we find that the observer's proper time t measured from the point $u=0$, is given by $u = \tanh\frac{t}{R}$, so that $t=-\infty$ corresponds to $u=-1$. Thus the general equation of a world-line intersecting the observer's at time $t=-\infty$ is

$$\frac{x}{a}=\frac{y}{b}=\frac{z}{c}=u+1.$$

For greater generality, however, we consider the world-line

$$\frac{x}{a}=\frac{y}{b}=\frac{z}{c}=u-U \quad , \qquad (79)$$

intersecting the observer's at $u=U$, where U is probably not far from -1, and is anyhow negative.

Suppose that the light which reaches the observer at the origin left the star at the point $\left(a(u_1-U), b(u_1-U), c(u_1-U), u_1\right)$. Then the interval between this point and (0000) is zero.

The condition for this is given by substitution in (57):

$$u_1^2(d^2-1) - 2d^2U . u_1 + d^2U^2 = 0 \qquad (80)$$

where $d=\sqrt{a^2+b^2+c^2}$.

Hence: $u_1 = \dfrac{Ud}{d\pm 1}$.

One value of u_1 must be positive and one negative; we must take the negative solution: $u_1 = \dfrac{Ud}{d+1}$. (cf. page 71)

D and D_0 can be calculated from (60). We find:

$$\cos^2 \frac{D_0}{R} = \frac{1-d^2}{d^2U^2-d^2+1} \quad , \quad \cos^2\frac{D}{R} = \frac{1-u_1^2}{1+d^2(u_1-U)^2-u_1^2} = 1-u_1^2, \text{ by } (80).$$

From (70) and (72):

$$\frac{\Delta}{R} = \frac{-Ud}{\sqrt{d^2U^2-d^2+1}} = -Ud - \tfrac{1}{2}Ud^3(1-U^2) - \cdots \qquad (81)$$

$$\frac{\delta\lambda}{\lambda} = \frac{-Ud}{(-u_1)\sqrt{d^2U^2-d^2+1}} - 1 = \frac{d+1}{\sqrt{d^2U^2-d^2+1}} - 1 = d + \tfrac{1}{2}d^2(d+1)(1-U^2) + \cdots \qquad (82)$$

We see that to a first approximation (neglecting $1-U^2$)

$$\frac{\delta\lambda}{\lambda} = -\frac{1}{U}\cdot\frac{\Delta}{R}. \qquad (83)$$

Comparing the observed correlation (see page 10), namely:

$$\frac{\delta\lambda}{\lambda} = \frac{558 \text{ Km/sec}}{\text{velocity of light}} \cdot \frac{\Delta}{10^6 \text{ parsecs}}, \qquad (84)$$

and taking $U=-1$, we find:

$R = 5.4 \times 10^8$ parsecs $= 1.7 \times 10^9$ light-years.

To the order of approximation which Tolman used, this agrees with his result, a slight discrepancy arising only from the slightly different data which we have used, not from the corrected definition of distance.

By eliminating d from (81) and (82) before expanding in powers of $(1-U^2)$, we obtain in place of (83) the closer approximation: $\dfrac{\delta\lambda}{\lambda} = -\dfrac{1}{U}\cdot\dfrac{\Delta}{R} + \dfrac{1}{2}\cdot\dfrac{1-U^2}{U^2}\cdot\dfrac{\Delta^2}{R^2} + \cdots$ (85)

There is reason to hope that a more accurate correlation than (84), involving a small term in Δ^2, will ultimately be established. If the universe were exactly of de Sitter's form, (85) would then give, on the hypothesis we have adopted U as well as $\mathcal{R}$; and hence we should know the "age of creation" $\mathcal{R}$ arc tanh U.

The hypothesis which we have been considering (that the world-lines of all the nebulae radiate approximately from a world-point in the remote past) seems at first sight very similar to the assumption of §9 that the nebulae are on the whole stationary in a "cosmic" system of coordinates. That the two assumptions are not equivalent is seen from equations (19), in which, if x_1, x_2, x_3 are constant, x,y,z are also constant and not functions of u as in (79).

Let us consider this second hypothesis, and see what information it leads to as to the size of the universe. In view of Eddington's remarks quoted on page 20, it is evident that a hypothesis like this, having no reference to the remote past, is to be preferred.

We have then to consider world-lines of the form

$$x = a, \ y = b, \ z = c,$$

where a,b,c, are constants. As before we shall take the observer's world-line to be x=y=z=0, so that in the cosmic coordinates (62) he is at rest at the spatial origin. Let the star be at (a,b,c,u_1) and the observer at $(0,0,0,u_0)$ at the instants of emission and reception of light. (We cannot take the observer to be at $(0,0,0,0)$, because now his u-coordinate is given by $u_0 = \tanh \frac{t}{R}$ in terms of his cosmic coordinate t which occurs in (62), determining how far the spatial expansion of the universe has proceeded.)

For the star to be seen by the observer, u_1 must be the smaller solution of:

$$(1-u_0^2)(1+d^2-u_1^2) = (1-u_0 u_1)^2, \quad \text{where } d = \sqrt{a^2+b^2+c^2}.$$

i.e. $u_1 = u_0 - d\sqrt{1-u_0^2}$

$$\frac{\Delta^2}{R^2} = \sin^2\frac{D_0}{R} = 1 - \frac{1-(d^2+1)u_0^2}{(d^2+1)(1-u_0^2)} = \frac{d^2}{(d^2+1)(1-u_0^2)},$$

whence $$\frac{1}{d^2+1} = 1 - (1-u_0^2)\frac{\Delta^2}{R^2}. \quad (85)$$

$$\sin^2\frac{D}{R} = 1 - \frac{1-u_1^2}{1+d^2-u_1^2} = \frac{d^2}{1+d^2-u_1^2} = \frac{d^2}{\left(u_0 d + \sqrt{1-u_0^2}\right)^2}.$$

$$\frac{\delta\lambda}{\lambda} = \frac{\sin\frac{D_0}{R}}{\sin\frac{D}{R}} - 1 = \frac{u_0 d + \sqrt{1-u_0^2}}{\sqrt{d^2+1}\sqrt{1-u_0^2}} - 1 = u_0\frac{\Delta}{R} + \frac{1}{\sqrt{d^2+1}} - 1$$

$$= u_0\frac{\Delta}{R} + \left[1 - (1-u_0^2)\frac{\Delta^2}{R^2}\right]^{\frac{1}{2}} - 1, \text{ by (85)},$$

$$= u_0\frac{\Delta}{R} - \frac{1}{2}(1-u_0^2)\frac{\Delta^2}{R^2} + \cdots$$

$$= \frac{\Delta}{R}\tanh\frac{t}{R} - \frac{\Delta^2}{2R^2}\operatorname{sech}^2\frac{t}{R} + \cdots \quad (86)$$

Thus, to a first approximation, $\frac{\delta\lambda}{\lambda}$ is proportional to Δ; but we should require to know the magnitude of the second term (in Δ^2) in the correlation (84), before we could effectively compare the two results. We should then be able to determine:

(i) the 4-dimensional radius $\mathcal{R}$;

(ii) the observer's coordinate t which determines how far the spatial expansion of the universe has proceeded;

(iii) the spatial radius of curvature $R = \mathcal{R} \cosh \frac{t}{\mathcal{R}}$.
[See note (2), page 8.]

§18. CONCLUSION.

I have applied my formulae for astronomical distance to the de Sitter world to show how they work. In view of the probability (see page 19) that the real universe approximates to the de Sitter world as regards the future and the recent past, but not as regards the remote past, the numerical results on page 83 are not reliable. But the formula (85) should give a reliable approximation when the more accurate form of (84) is established. So far as I am aware, the formula (86) has not been given before.

As I have already indicated, I hope in a future investigation to apply my formulae for distance to the Expanding Universe, and see whether any corrections are needed on the numerical results which have been established.

[The End.

Supplement

On Errors in Determinants

by

I. M. H. Etherington, B.A.
University of Edinburgh
Mathematical Dept.
16 Chambers Street
Edinburgh.

This paper has been accepted for publication in the Proceedings of the Edinburgh Mathematical Society

On Errors in Determinants.

On Errors in Determinants.

§1. Introduction.

Care is needed in dealing with determinants whose elements are subject to experimental error, particularly when a determinant itself is small compared with its first minors. For, as these examples show, a relatively tiny error in one element may be responsible for a large error in the determinant:

$$\begin{vmatrix} -73 & 78 & 24 \\ 92 & 66 & 25 \\ -80 & 37 & 10 \end{vmatrix} = 1 \;; \qquad \begin{vmatrix} -73 & 78 & 24 \\ 92 & 66 & 25 \\ -80 & 37 & 10\cdot01 \end{vmatrix} = -118\cdot94 .$$

Consider a determinant Δ, of order n, with elements a_i $(i = 1 \dots n^2)$, & let A_i, A_{ij}, A_{ijk}, ... denote the first, second, third, ... minors corresponding to the elements indicated by the suffixes. Let the actual error in each element a_i be e_i, and let the resultant error in Δ be E. In §§2-4, I shall find expressions for (i) E in terms of e_i; (ii) the ~~maximum~~ range of E in terms of the ranges of e_i; (iii) the probability distribution $P(E)$ of E in terms of the probability distributions $p_i(e_i)$ of e_i, assumed independent and (for simplicity) symmetrical. In §§5-7, I shall investigate corresponding results for the quotient of two determinants which are identical except for one row or column; such quotients are of importance in practical work, occurring in the solution of a set of simultaneous linear algebraic equations. In §8 the method is applied to an arbitrary function whose arguments are subject to error.

To apply the formulae ~~*~~ to numerical determinants, it is necessary for first approximations to calculate the complete set of first minors; the second, third, etc, minors are required for closer approximations. The calculation of these minors is very laborious in the case of a determinant of large order. It may be pointed out, however, that all the first minors can be found in the course of evaluating the determinant by Dr. Aitken's method, explained in the preceding paper. In any case, the calculations are facilitated by a calculating machine.

§2. Actual Error in a Determinant.

~~Let A_i, A_{ij}, A_{ijk}, … denote the first~~

An error e_i in the term a_i will cause an error $e_i A_i$ in Δ; errors e_i, e_j in the terms a_i, a_j will together cause an error $(e_i A_i + e_j A_j + e_i e_j A_{ij})$ in Δ; and so on. Hence in general:

$$E = \sum e_i A_i + \sum e_i e_j A_{ij} + \sum e_i e_j e_k A_{ijk} + \cdots, \text{ to } n \text{ summations}, \qquad (1)$$

the suffixes in each summation running from 1 to n^2.

(This result follows at once from Taylor's Theorem, A_i, A_{ij}, A_{ijk}, … being partial derivatives of Δ.) It is to be observed that none of the errors occur squared or to higher powers in (1).

§3. Range of Error.

If $|e_i| \leq \epsilon_i$, we deduce from (1):

$$|E| \leq \sum \epsilon_i |A_i| + \sum \epsilon_i \epsilon_j |A_{ij}| + \sum \epsilon_i \epsilon_j \epsilon_k |A_{ijk}| + \cdots, \qquad (2)$$

vertical bars denoting absolute values.

If we put $\epsilon_i = \epsilon$ and neglect ϵ^2, (2) becomes:

$$|E| \leq \epsilon \sum |A_i|. \qquad (3)$$

Thus $\sum |A_i|$ may be regarded as a measure of the sensitivity of a determinant to small errors whose squares may be neglected.

As an application of these formulae, consider the first determinant in §1, & suppose that the range of error in each element is $\pm\frac{1}{2}$. We find $\Sigma A_i = 33099$, $\Sigma|A_{ij}| = 2\,\Sigma|a_i| = 970$, and (for any third order determinant) $\Sigma|A_{ijk}| = 6$. Hence ③ gives ± 16550 as a first approximation to the range of error in Δ. More accurately, ④ gives:

$$|E| \leq \frac{1}{2}\cdot 33099 + \frac{1}{4}\cdot 950 + \frac{1}{8}\cdot 6 \quad < 16793.$$

If we suppose that the determinant has occurred in some practical calculation, & that the elements as given are only rough first approximations (with $\epsilon = \frac{1}{2}$), then the important question arises: to what further degree of accuracy η should the elements be evaluated in order that the error in the determinant may not exceed a given limit, say for example 0.2 ?

We have in general:

$$|E| \leq \eta\,\Sigma|B_i| + \eta^2\,\Sigma|B_{ij}| + \cdots,$$

where B_i, B_{ij}, ... are the <u>correct</u> values of the minors to which A_i, A_{ij}, ... are approximations. Now

$$|A_i - B_i| = \text{error in minor of } a_i \text{ in the first approximation}$$
$$\leq \epsilon \sum_j |A_{ij}| + \epsilon^2 \sum_{jk} |A_{ijk}| + \cdots$$
$$\therefore |B_i| \leq |A_i| + \epsilon \sum_j |A_{ij}| + \epsilon^2 \sum_{jk} |A_{ijk}| + \cdots.$$

Summing for $i = 1 \dots n^2$,

$$\sum |B_i| \leqslant \sum |A_i| + \epsilon \sum |A_{ij}| + \epsilon^2 \sum |A_{ijk}| + \cdots .$$

Similarly,

$$\sum |B_{ij}| \leqslant \sum |A_{ij}| + \epsilon \sum |A_{ijk}| + \cdots \quad , \quad \text{etc.}$$

$$\therefore \ |E| \leqslant \eta \left[\sum |A_i| + \epsilon \sum |A_{ij}| + \epsilon^2 \sum |A_{ijk}| + \cdots \right]$$
$$+ \eta^2 \left[\sum |A_{ij}| + \epsilon \sum |A_{ijk}| + \cdots \right] + \cdots .$$

This formula gives the range of error of a determinant when the elements are correct to within $\pm \eta$, in terms of the minors in a first approximation in which the elements are correct to within $\pm \epsilon$.

Let us apply this to the determinant already considered, for which, with $\epsilon = \frac{1}{2}$,

$$\sum |A_i| + \epsilon \sum |A_{ij}| + \epsilon^2 \sum |A_{ijk}| = 33585\tfrac{1}{2} .$$

Putting $\eta = \cdot 000005$, we find $|E| < 0{\cdot}17$.

Thus, in order that the error in Δ should not exceed $0{\cdot}2$, it is sufficient to evaluate each element to 5 decimal places.

§4. Probability Distribution.

Denote the probability distribution of e_i by $p_i(e_i)$, and the r^{th} moment of this function by

$$m_{ri} \equiv \int_{-\infty}^{\infty} e_i^r \, p_i(e_i) \, de_i \quad . \qquad (4)$$

The functions p_i being given, we can take it that these moments m_{ri} are known; in terms of them we can calculate the r^{th} moment M_r of $P(E)$.

We have in fact:

$$M_r \equiv \int_{-\infty}^{\infty} E^r P(E) \, dE = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} E^r . \prod_1^{n^2} \left[p_i(e_i) \, de_i \right] . \qquad (5)$$

On the right hand side of (5), E^r is to be interpreted as a function of e_i, by means of (1). Consequently, for any value of r, we can expand E^r and integrate term by term, making use of (4), & of the conditions $\int_{-\infty}^{\infty} p_i \, de_i = 1$.

For example,

$$E^2 = \Sigma A_i^2 e_i^2 + 2 \Sigma A_i A_j e_i e_j + 2 \Sigma A_i A_{ij} e_i^2 e_j$$
$$+ 2 \Sigma A_i A_{jk} e_i e_j e_k + \cdots .$$

On integration this gives:

$$M_2 = \Sigma A_i^2 m_{2i} + 2 \Sigma A_i A_j m_{1i} m_{1j} + 2 \Sigma A_i A_{ij} m_{2i} m_{1j}$$
$$+ 2 \Sigma A_i A_{jk} m_{1i} m_{1j} m_{1k} + \cdots .$$

The given functions have been assumed ~~to~~ to be symmetrical,

i.e. $p_i(e_i) = p_i(-e_i)$;

consequently the odd moments m_{1i}, m_{3i}, ... all vanish

So we need only take account of those terms of E^r which contain only even powers of the errors e_i; all other terms vanish on integration. We then find :—

$M_1 = 0$; (6)

$M_2 = \Sigma A_i^2 m_{2i} + \Sigma A_{ij}^2 m_{2i} m_{2j} + \dots$ to n summations; (7)

$M_3 = 6 \Sigma A_i A_j A_{ij} m_{2i} m_{2j} + \dots$;

$M_4 = \Sigma A_i^4 m_{4i} + 6 \Sigma A_i^2 A_j^2 m_{2i} m_{2j} + \dots$;

Etc.

Each of these expressions has only a finite number of terms, but they rapidly become very complicated. However, if the original errors are known to be sufficiently small, i.e. if $p_i(e_i)$ are practically zero except when e_i are small, then the moments m_{ri} are also small quantities; we can therefore approximate to M_r successfully with only a few terms, & the succeeding moments M_r will rapidly diminish in order of magnitude.

With a change of notation (7) can be written:

$$S^2 = \Sigma A_i^2 \sigma_i^2 + \Sigma A_{ij}^2 \sigma_i^2 \sigma_j^2 + \cdots \qquad (8)$$

giving the standard deviation S $(=\sqrt{M_2})$ of the determinant in terms of the standard deviations σ_i $(=\sqrt{m_{2i}})$ of the elements.

Having thus found $M_1, M_2, M_3, \cdots$, we can approximate to $P(E)$ in one of the usual ways. For instance, if we assume that the errors e_i follow the normal law of distribution $p_i(e_i) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{e_i^2}{2\sigma_i^2}}$, (or, more generally, that they follow a symmetrical law of Charlier's Type A,) then the law for the determinant must necessarily be of Charlier's Type A. We can thus assume

$$P(E) = \left[1 + A\frac{d^3}{dE^3} + B\frac{d^4}{dE^4} + \cdots\right] \frac{1}{S\sqrt{2\pi}} e^{-\frac{E^2}{2S^2}}.$$

S is given by equation (8) and the other constants by: $A = -\frac{M_3}{6}$, $B = \frac{M_4}{24} - \frac{S^4}{8}$, etc., these results being found by comparing the moments of the function with $M_2, M_3, \cdots$.

To a first approximation,

$$S^2 = \sum A_i^2\sigma_i^2 + \sum A_i^2 A_j^2 \sigma_i^2 \sigma_j^2 ;$$

$$A = -\sum A_i A_j A_{ij} \sigma_i^2 \sigma_j^2 ;$$

$$24B = \sum A_i^4 m_{4i} + 6\sum A_i^2 A_j^2 \sigma_i^2 \sigma_j^2 - 3\left(\sum A_i^4 \sigma_i^4 + 2\sum A_i^2 A_j^2 \sigma_i^2 \sigma_j^2\right)$$

$= 0$, since for <u>normal</u> distributions $m_{4i} = 3\sigma_i^4$.

For practical purposes, a more important case arises when the elements have been calculated correct to x places of decimals, the last figure being forced. (i.e. the last digit retained is increased by 1 when the first digit not retained is 5, 6, 7, 8, or 9.) We may take it that all errors between $\pm\frac{1}{2}10^{-x}$ are equally likely, & that none exceed these limits. Remembering that $\int_{-\infty}^{\infty} p(e)\,de$ must be unity, we have:

$$p(e) = 10^x, \quad \text{if } -\tfrac{1}{2}\cdot 10^{-x} \leq e \leq \tfrac{1}{2}10^{-x},$$
$$= 0, \quad \text{otherwise};$$

the same law applying to all the elements.

Then: $m_1 = m_3 = \cdots = 0$;

$$m_2 = \int_{-\frac{1}{2}10^{-x}}^{\frac{1}{2}10^{-x}} e^2 . 10^x \, de = \frac{1}{12 . 10^{2x}} ;$$

$$m_4 = \int_{-\frac{1}{2}10^{-x}}^{\frac{1}{2}10^{-x}} e^4 . 10^x . de = \frac{1}{80 . 10^{4x}} ;$$

Neglecting terms involving 10^{-6x},

$$M_1 = 0 ;$$

$$M_2 = \frac{\Sigma A_i^2}{12 . 10^{2x}} + \frac{\Sigma A_{ij}^2}{144 . 10^{4x}} ;$$

$$M_3 = \frac{\Sigma A_i^4}{80 . 10^{4x}} + \frac{\Sigma A_i^2 A_j^2}{24 . 10^{4x}} ,$$

Hence, to this order of approximation, we find:

$$P(E) = \left[1 - \frac{\Sigma A_i A_j A_{ij}}{144 . 10^{4x}} . \frac{d^3}{dE^3} - \frac{\Sigma A_i^4}{288 . 10^{5x}} . \frac{d^4}{dE^4} \right] \frac{1}{S\sqrt{2\pi}} . e^{-\frac{E^2}{2S^2}} \qquad (9)$$

where $$S^2 = \frac{\Sigma A_i^2}{12 . 10^{2x}} + \frac{\Sigma A_{ij}^2}{144 . 10^{4x}} . \qquad (10)$$

As an application of these formulae, consider the determinant:

$$\begin{vmatrix} \sqrt{7} & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & \sqrt{11} & \sqrt{3} \\ \sqrt{3} & \sqrt{5} & \sqrt{6} \end{vmatrix},$$

whose value to five places is 6.93899. Evaluating each element to two places, we get

$$\begin{vmatrix} 2.65 & 1.41 & 1.41 \\ 1.41 & 3.32 & 1.73 \\ 1.73 & 2.24 & 2.45 \end{vmatrix},$$

whose value to five places is 6.98983.

For the latter determinant we find that (correct to the number of places given in each case):

$$\Sigma |A_i| = 26.81$$

$$\Sigma A_i^2 = 111.9$$

$$\Sigma A_{ij}^2 = 78$$

$$\Sigma A_i A_j A_{ij} = 287$$

$$\Sigma A_i^4 = 1078$$

$$\epsilon = .005 .$$

By (3), the maximum range of error is $\pm .135$.
By (10), the standard deviation (to a first approximation) is .031.
These results may be compared with the actual error .051.
The second approximation to the probability distribution is given by

$$S = .0314, \quad A = -.0002, \quad B = .0004 .$$

§5. Error in Quotient of Determinants.

Consider the quotient $\frac{\Delta_1}{\Delta_2}$ of the determinants*

$$\Delta_1 = \left| a_i \; a_{i1} \; a_{i2} \; \cdots\cdots \; a_{i\,\overline{n-1}} \right| , \quad (i = 1 \ldots n)$$

$$\Delta_2 = \left| b_i \; a_{i1} \; a_{i2} \; \cdots\cdots \; a_{i\,\overline{n-1}} \right| .$$

Suppose that the elements a_i, b_i, a_{ij} are subject to errors e_i, f_i, e_{ij}; and that the resultant errors in Δ_1, Δ_2, $\frac{\Delta_1}{\Delta_2}$ are E_1, E_2, E.

Form the determinant

$$\Delta_3 = \left| c_i \; a_{i1} a_{i2} \; \cdots \; a_{i\,\overline{n-1}} \right| ,$$

where $c_i = a_i - X b_i$,

X = the calculated value of $\frac{\Delta_1}{\Delta_2}$;

so that Δ_3 = the calculated value of $\Delta_1 - X\Delta_2 = 0$. (11)

(But the correct value of $\Delta_1 - X\Delta_2 \neq 0$.)

---

* i.e. the determinants of which these are the i^{th} rows. The determinants are identical except in their first columns. As pointed out in §1, such quotients occur in the solution of a set of n simultaneous linear equations.

Let g_i = error in c_i $= e_i - X f_i$,

E_3 = error in $\Delta_3 = E_1 - XE_2$,

A_i = minor of a_i in Δ_1

= minor of b_i in Δ_2

= minor of c_i in Δ_3 ,

A_{ij} = minor of a_{ij} in Δ_1 ,

B_{ij} = minor of a_{ij} in Δ_2 ,

C_{ij} = minor of a_{ij} in $\Delta_3 = A_{ij} - X B_{ij}$.

It follows from (11) that the minors of any one row of Δ_3 are proportional to the minors of the first row, i.e.

$$C_{ij} = \frac{A_i C_{1j}}{A_1} . \qquad (12)$$

Using (1), we have as first approximations:

$$E_2 = \sum f_i A_i + \sum e_{ij} B_{ij} ,$$

$$E_3 = \sum g_i A_i + \sum e_{ij} C_{ij} .$$

Correct to the second order,

$$E = \frac{\Delta_1 + E_1}{\Delta_2 + E_2} - \frac{\Delta_1}{\Delta_2}$$

$$= \frac{\Delta_2 E_1 - \Delta_1 E_2}{\Delta_2 (\Delta_2 + E_2)}$$

$$= \frac{E_1 - XE_2}{\Delta_2} \left(1 + \frac{E_2}{\Delta_2}\right)^{-1}$$

$$= \frac{E_3}{\Delta_2} \left(1 - \frac{E_2}{\Delta_2}\right) .$$

Thus, to the first order,

$$E = \frac{1}{\Delta_2}\left[\Sigma(e_i - Xf_i)A_i + \Sigma e_{ij} C_{ij}\right]. \qquad (13)$$

To the second order,

$$E = \frac{E_3}{\Delta_2} - \frac{1}{\Delta_2^2}\left[\Sigma f_i A_i + \Sigma e_{ij} B_{ij}\right]\left[\Sigma(e_i - Xf_i)A_i + \Sigma e_{ij} C_{ij}\right], \qquad (14)$$

$$E^2 = \frac{1}{\Delta_2^2}\left[\Sigma(e_i - Xf_i)A_i + \Sigma e_{ij} C_{ij}\right]^2. \qquad (15)$$

§6. Range of Error in Quotient of Determinants.

Suppose $|e_i| \leq \epsilon_i$, $|f_i| \leq \mathbf{f}_i$, $|e_{ij}| \leq \epsilon_{ij}$.

From (13), the range of E is given by:

$$|E| \leq \frac{1}{|\Delta_2|}\cdot\left[\Sigma(\epsilon_i + |X|\,\mathbf{f}_i).|A_i| + \Sigma\,\epsilon_{ij}.|C_{ij}|\right].$$

If the given ranges are all equal, i.e. $\epsilon_i = \mathbf{f}_i = \epsilon_{ij} = \epsilon$, we have, using (12):

$$|E| \leq \frac{\epsilon}{|\Delta_2|}\cdot\left[\Sigma(1+|X|).|A_i| + \Sigma\left|\frac{A_i C_{ij}}{A_1}\right|\right]$$

$$\leq \frac{\epsilon}{|\Delta_2 A_1|}\cdot\Sigma|A_i|\cdot\left[(1+|X|)A_1 + \Sigma|C_{1j}|\right]. \qquad (16)$$

This can be expressed in the form $|E| \leq \epsilon\cdot\frac{1+|X|}{|\Delta_2 A_1|}\cdot S_1 S_2$,

where S_1 = sum of absolute values of 1st minors of 1st column of D,
S_2 = " " " " " row of D.

D being the determinant $\left|\frac{a_i - Xb_i}{1+|X|}\; a_{i1}\; a_{i2} \cdots a_{i\,n-1}\right|$.

§7. Probability Distribution for Quotient of Determinants.

Let the standard deviations of e_i, f_i, e_{ij} be σ_i, τ_i, σ_{ij}, it being assumed that the first moments are zero. Let M_1, M_2, M_3 ... be the moments of the probability distribution $P(E)$ of E. Using (14) & (15), proceeding as in §4, & remembering that in accordance with (6) the first moment of E_3 is zero, we get, correct to the second order in the deviations:

$$M_1 = \frac{1}{\Delta_2^2}\left[X.\sum \tau_i^2 A_i^2 - \sum \sigma_{ij}^2 B_{ij} C_{ij}\right],$$

$$M_2 = \frac{1}{\Delta_2^2}\left[\sum(\sigma_i^2 + X^2\tau_i^2)A_i^2 + \sum \sigma_{ij}^2 C_{ij}^2\right],$$

$$M_3 = M_4 = \cdots = 0 .$$

Thus, assuming that the required function is of Charlier's Type A, we have as a first approximation:

$$P(E) = \frac{1}{S\sqrt{2\pi}}\, e^{-\frac{(E-a)^2}{2S^2}}$$

where

$$a = M_1 = \frac{1}{\Delta_2^2}\left[X \sum \tau_i^2 A_i^2 - \sum \sigma_{ij}^2 B_{ij} C_{ij}\right],$$

$$S^2 = M_2 - M_1^2 = \frac{1}{\Delta_2^2}\left[\sum(\sigma_i^2 + X^2\tau_i^2)A_i^2 + \sum \sigma_{ij}^2 C_{ij}^2\right].$$

The second approximation to $P(E)$ involves many complicated summations.

If the given standard deviations are all equal, i.e. $\sigma_i = \tau_i = \sigma_{ij} = \sigma$, we have, using (12):

$$a = \frac{\sigma^2}{\Delta_2^2}\left[X \cdot \Sigma A_i^2 - \frac{1}{A_1}\,\Sigma B_{ij} A_i C_{ij}\right], \qquad (17)$$

$$S^2 = \frac{\sigma^2}{\Delta_2^2 A_1^2} \cdot \Sigma A_i^2 \cdot \left[(1+X^2)A_1^2 + \Sigma C_{ij}^2\right]. \qquad (18)$$

As in § 6, we ~~state~~ note that ~~the~~ (18) can be written:

$$S^2 = \sigma^2 \frac{1+X^2}{\Delta_2^2 A_1^2} S_3 S_4 ,$$

where S_3 = sum of squares of first minors of first column of D',
S_4 = " " " " " row of D',
D' being the determinant $\left| \frac{a_i - X b_i}{\sqrt{1+X^2}} \; a_{i1} \; a_{i2} \ldots\ldots a_{i\,n-1} \right|$.

As an application of these formulae, consider the value of X found from the equations:

$$X\sqrt{7} + Y\sqrt{2} + Z\sqrt{2} = \sqrt{17}$$
$$X\sqrt{3} + Y\sqrt{11} + Z\sqrt{3} = -\sqrt{2}$$
$$X\sqrt{2} + Y\sqrt{5} + Z\sqrt{6} = -\sqrt{3} .$$

To four places, the value is 3·1468.

Let us however approximate by evaluating each coefficient to two places. We then have:

$$\Delta_1 = \begin{vmatrix} 4\cdot12 & 1\cdot41 & 1\cdot41 \\ -1\cdot41 & 3\cdot32 & 1\cdot73 \\ -1\cdot73 & 2\cdot24 & 2\cdot45 \end{vmatrix} = 21\cdot842244,$$

$$\Delta_2 = \begin{vmatrix} 2\cdot65 & 1\cdot41 & 1\cdot41 \\ 1\cdot41 & 3\cdot32 & 1\cdot73 \\ 1\cdot73 & 2\cdot24 & 2\cdot45 \end{vmatrix} = 6\cdot989832 .$$

The following results are correct to the number of places given in each case.

$$X = \frac{\Delta_1}{\Delta_2} = 3\cdot1249 .$$

$$\Delta_3 = \begin{vmatrix} -4\cdot16 & 1\cdot41 & 1\cdot41 \\ -5\cdot82 & 3\cdot32 & 1\cdot73 \\ -7\cdot14 & 2\cdot24 & 2\cdot45 \end{vmatrix} = \cdot01$$

$$A_1 = 4\cdot26$$

$$\Sigma |A_i| = 6\cdot8 , \quad \Sigma A_i^2 = 23\cdot25, \quad \Sigma B_{ij} A_i C_{1j} = -264 .$$

$$(1+|x|).|A_1| + \Sigma |C_{1j}| = 30$$

$$X \Sigma A_i^2 - \frac{1}{A_1} \Sigma B_{ij} A_i C_{1j} = 134$$

$$(1+X^2)A_1^2 + \Sigma C_{1j}^2 = 313$$

$$\epsilon = \cdot005 , \quad \sigma^2 = \frac{1}{12.10^4} .$$

From (16), (17) & (18) we deduce:

Range of error $= \pm .035$,

$a = .0002$,

$\delta = .008$.

The actual error is $.022$.

§8. Error in an Arbitrary Function.

The method of §4 may also be used to approximate to the probability distribution of error in an arbitrary function F (instead of the determinant Δ) of the quantities a_i. Equation (1) is then formally the same, but $\left(\text{since } F_{ii} \equiv \frac{\partial^2 F}{\partial a_i^2} \neq 0\right)$ it includes terms with higher powers of the errors e_i.

The following are corresponding results (suffixes of F indicating partial derivatives). –

$$M_1 = \tfrac{1}{2} \sum F_{ii}\, m_{2i} + \tfrac{1}{24} \sum F_{iiii}\, m_{4i} + \tfrac{1}{4} \sum F_{iijj}\, m_{2i}\, m_{2j} + \cdots;$$

$$M_2 = \sum F_i^2 m_{2i} + \tfrac{1}{4} \sum F_{ii}^2 m_{4i} + \sum \left(F_{ii} F_{jj} + F_{ij}^2\right) m_{2i}\, m_{2j} + \cdots;$$

$$M_3 = \tfrac{3}{2} \sum F_i^2 F_{ii}\, m_{4i} + \left(\tfrac{3}{2} \sum F_i^2 F_{jj} + 6 F_i F_j F_{ij}\right) m_{2i}\, m_{2j} + \cdots;$$

$$M_4 = \sum F_i^4 m_{4i} + 6 \sum F_i^2 F_j^2 m_{2i}\, m_{2j} + \cdots. \quad (i \neq j)$$

If the given laws of distribution are normal, the required function will be of the form

$$P(E) = \left[1 + A\frac{d^3}{dE^3} + B\frac{d^4}{dE^4} + \ldots\ldots\right]\frac{1}{S\sqrt{2\pi}}\,e^{-\frac{(E-a)^2}{2S^2}}.$$

The constants are given by:

$$a = M_1,\quad S^2 = M_2 - M_1^2,\quad A = -\frac{M_3}{6} + \frac{M_1M_2}{2} - \frac{M_1^3}{3},\quad B = \frac{M_4}{24} + \frac{M_1^4}{12} - \frac{M_2^2}{8} + M_1A.$$

To a first approximation:

$$a = \tfrac{1}{2}\Sigma F_{ii}\sigma_i^2;\quad S^2 = \Sigma F_i^2\sigma_i^2;\quad A, B, \ldots = 0.$$

To a second approximation:

$$a = \tfrac{1}{2}\Sigma F_{ii}\sigma_i^2 + \tfrac{1}{8}\Sigma F_{iiii}\sigma_i^4 + \tfrac{1}{4}\Sigma F_{iijj}\sigma_i^2\sigma_j^2;$$

$$S^2 = \Sigma F_i^2\sigma_i^2 + \tfrac{1}{2}\Sigma F_{ii}^2\sigma_i^4 + \Sigma F_{ij}^2\sigma_i^2\sigma_j^2;$$

$$A = -\tfrac{1}{2}\Sigma F_{ii}F_i^2\sigma_i^4 - \Sigma F_{ij}F_iF_j\sigma_i^2\sigma_j^2;$$

$$B, \ldots = 0.$$

If the arguments of F have been calculated correct to x decimal places, the last figure being forced, then the second approximation is given by:

$$a = \frac{\Sigma F_{ii}}{24.10^{2x}} + \frac{\Sigma F_{iiii}}{1920.10^{4x}} + \frac{\Sigma F_{iijj}}{576.10^{4x}};$$

$$S^2 = \frac{\Sigma F_i^2}{12.10^{2x}} + \frac{\Sigma F_{ii}^2}{720.10^{4x}} + \frac{\Sigma F_{ii}F_{jj}}{288.10^{4x}} + \frac{\Sigma F_{ij}^2}{144.10^{4x}};$$

$$A = -\frac{\Sigma F_{ii}F_i^2}{720.10^{4x}} - \frac{\Sigma F_iF_jF_{ij}}{144.10^{4x}};$$

$$B = -\frac{\Sigma F_i^4}{2880.10^{4x}}.$$

§9. Acknowledgment. My best thanks are due to Dr. A. C. Aitken, whose suggestions gave rise to this investigation; & to Mr. G. Lidstone for his valuable criticisms.