Mr. Galton then read the following paper by the Rev. H. W. Watson and himself:

On the PROBABILITY *of the* EXTINCTION *of* FAMILIES. By the Rev. H. W. WATSON. With PREFATORY REMARKS, by FRANCIS GALTON, F.R.S.

THE decay of the families of men who occupied conspicuous positions in past times has been a subject of frequent remark, and has given rise to various conjectures. It is not only the families of men of genius or those of the aristocracy who tend to perish, but it is those of all with whom history deals, in any way, even of such men as the burgesses of towns, concerning whom Mr. Doubleday has inquired and written. The instances are very numerous in which surnames that were once common have since become scarce or have wholly disappeared. The tendency is universal, and, in explanation of it, the conclusion has been hastily drawn that a rise in physical comfort and intellectual capacity is necessarily accompanied by diminution in "fertility" —using that phrase in its widest sense and reckoning abstinence from marriage as sterility. If that conclusion be true, our population is chiefly maintained though the "proletariat," and thus a large element of degradation is inseparably connected with those other elements which tend to ameliorate the race. On the other hand, M. Alphonse De Candolle has directed attention to the fact that, by the ordinary law of chances, a large proportion of families are continually dying out, and it evidently follows that, until we know what that proportion is, we cannot estimate whether any observed diminution of surnames among the families whose history we can trace, is or is not a sign of their diminished "fertility." I give extracts from M. De Candolle's work in a foot-note,* and may add that, although I have not hitherto published anything on the matter, I took considerable pains some years ago to obtain numerical results in respect to this very problem. I made certain very simple, but not very inaccurate, suppositions, concerning average fertility, and I worked to the nearest integer, starting with 10,000 persons, but the computation became intolerably tedious after a few steps, and I had to abandon it. More recently, having first privately applied in vain to some

* "Au milieu des renseignements précis et des opinions très-sensées de MM. Benoiston de Châteauneuf, Galton, et autres statisticiens, je n'ai pas rencontré la réflexion bien importante qu'ils auraient dû faire de l'extinction *inévitable* des noms defamille. Évidemment tous les noms doivent s'éteindre ....Un mathématicien pourrait calculer comment la réduction des noms ou titres aurait lieu, d'après la probabilité des naissances toutes féminines ou toutes masculines ou mélangées et la probabilité d'absence de naissances dans un couple quelconque," etc.—Alphonse de Candolle, "Histoire des Sciences et des Savants," 1873.

mathematicians, I put the problem into a shape suited to mathematical treatment, and proposed it in the pages of a well-known mathematical periodical of a high class, the "Educational Times." It met with poor success at first, because the answer it received was from a correspondent who wholly failed to perceive its intricacy, and his results were totally erroneous. My friend the Rev. H. W. Watson then kindly, at my request, took the problem in hand, and published his first results in the above-mentioned periodical. These have since been considerably extended, and form the subject of the following paper. They do not give what can properly be called a general solution, but they do give certain general results. They show (1) how to compute, though with great labour, any special case; (2) a remarkably easy way of computing those special cases in which the law of fertility approximates to a certain specified form; and (3), how all surnames tend to disappear. I therefore feel sure that Mr. Watson's memoir will be of interest to the Anthropological Institute, and I beg to submit it to their notice, both for its intrinsic value and in hopes that other mathematicians may pursue the inquiry and attain still nearer to a complete solution of this very important problem.

The form in which I originally stated the problem is as follows. I purposely limited it in the hope that its solution might be more practicable if unnecessary generalities were excluded :—

A large nation, of whom we will only concern ourselves with the adult males, N in number, and who each bear separate surnames, colonise a district. Their law of population is such that, in each generation, a_0 per cent. of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 have two; and so on up to a_5 who have five. Find (1) what proportion of the surnames will have become extinct after r generations; and (2) how many instances there will be of the same surname being held by m persons.

Discussion of the problem by the Rev. H. W. WATSON.

Suppose that at any instant all the adult males of a large nation have different surnames, it is required to find how many of these surnames will have disappeared in a given number of generations upon any hypothesis, to be determined by statistical investigations, of the law of male population.

Let, therefore, a_0 be the percentage of males in any generation who have no sons reaching adult life, let a_1 be the percentage that have one such son, a_2 the percentage that have two, and so on up to a_q, the percentage that have q such sons, q being so large that it is not worth while to consider the chance of any man having more than q

adult sons—our first hypothesis will be that the numbers a_0, a_1, a_2, etc., remain the same in each succeeding generation. We shall also, in what follows, neglect the overlapping of generations—that is to say, we shall treat the problem as if all the sons born to any man in any generation came into being at one birth, and as if every man's sons were born and died at the same time. Of course it cannot be asserted that these assumptions are correct. Very probably accurate statistics would discover variations in the values of a_0, a_1, etc., as the nation progressed or retrograded; but it is not at all likely that this variation is so rapid as seriously to vitiate any general conclusions arrived at on the assumption of the values remaining the same through many successive generations. It is obvious also that the generations must overlap, and the neglect to take account of this fact is equivalent to saying, that at any given time we leave out of consideration those male descendants of any original ancestor who are more than a certain average number of generations removed from him, and compensate for this by giving credit for such male descendants, not yet come into being, as are not more than that same average number of generations removed from the original ancestors.

Let then $\frac{a_0}{100}$, $\frac{a_1}{100}$, $\frac{a_2}{100}$, etc., up to $\frac{a_q}{100}$, be denoted by the symbols t_0, t_1, t_2, etc., up to t_q, in other words, let t_0, t_1, etc., be the chances in the first and each succeeding generation of any individual man, in any generation, having no son, one son, two sons, and so on, who reach adult life. Let N be the original number of distinct surnames, and let ${}_rm_s$ be the fraction of N which indicates the number of such surnames with s representatives in the rth generation.

Now, if any surname have p representatives in any generation, it follows from the ordinary theory of chances that the chance of that same surname having s representatives in the next succeeding generation is the coefficient of x^s in the expansion of the multinomial

$$(t_0 + t_1x + t_2x^2 + \text{, etc. } + t_qx^q)^p$$

Let then the expression $t_0 + t_1x + t_2x^2 + \text{ etc. } + t_qx^q$ be represented by the symbol T.

Then since, by the assumption already made, the number of surnames with no representative in the $r-1$th generation is ${}_{r-1}m_0$ N, the number with one representative ${}_{r-1}m_1$. N, the number with two ${}_{r-1}m_2$. N and so on, it follows, from what we last stated, that the number of surnames with s representatives in the rth generation must be the coefficient of x^s in the expression

$$\left\{ {}_{r-1}m_0 + {}_{r-1}m_1\mathrm{T} + {}_{r-1}m_2\mathrm{T}^2 + \text{ etc. } + {}_{r-1}m_{qr-1}\mathrm{T}^{qr-1} \right\} \mathrm{N}$$

If, therefore, the coefficient of N in this expression be denoted by $f_r(x)$ it follows that ${}_{r-1}m_1$, ${}_{r-1}m_2$ and so on, are the coefficients of x, x^2 and so on, in the expression $f_{r-1}(x)$.

If, therefore, a series of functions be found such that

$$f_1(x) = t_0 + t_1x + \text{ etc. } + t_qx^q \text{ and } f_r(x) = f_{r-1}(t_0 + t_1x \text{ etc. } + t_qx^q)$$

then the proportional number of groups of surnames with s representatives in the rth generation will be the coefficient of x^s in $f_r(x)$ and the actual number of such surnames will be found by multiplying this coefficient by N. The number of surnames unrepresented or become extinct in the rth generation will be found by multiplying the term independent of x in $f_r(x)$ by the number N.

The determination, therefore, of the rapidity of extinction of surnames, when the statistical data, t_0, t_1, etc., are given, is reduced to the mechanical, but generally laborious process of successive substitution of $t_0 + t_1 x + t_2 x^2 +$ etc., for x in successively determined values of $f_r(x)$, and no further progress can be made with the problem until these statistical data are fixed; the following illustrations of the application of our formula are, however, not without interest.

(1) The very simplest case by which the formula can be illustrated is when $q=2$ and t_0, t_1, t_2 are each equal to $\frac{1}{3}$.

$$\text{Here } f_1(x)=\frac{1+x+x^2}{3} \qquad f_2(x)=\frac{1}{3}\left\{1+\frac{1}{3}(1+x+x^2)+\frac{1}{9}(1+x+x^2)^2\right\}^2$$

and so on.

Making the successive substitutions, we obtain

$$f_2(x)=\frac{1}{3}\left\{\frac{13}{9}+\frac{5x}{9}+\frac{6x^2}{9}+\frac{2x^3}{9}+\frac{x^4}{9}\right\}$$

$$f_3(x)=\frac{1249}{2187}+\frac{265x}{2187}+\frac{343x^2}{2187}+\frac{166x^3}{2187}+\frac{109x^4}{2187}+\frac{34x^5}{2187}+\frac{16x^6}{2187}+\frac{4x^7}{2187}+\frac{x^8}{2187}$$

$$f_4(x)=\cdot63183+\cdot08306x+\cdot10635x^2+\cdot07804x^3+\cdot06489x^4+\cdot05443x^5+\cdot01437x^6$$
$$+\cdot01692x^7+\cdot01144x^8+\cdot00367x^9+\cdot00104x^{10}+\cdot00015x^{11}+\cdot00005x^{12}$$
$$+\cdot00001x^{13}+\cdot00000x^{14}+\cdot00000x^{15}+\cdot00000x^{16}$$

and the constant term in $f_5(x)$ or ${}_5m_0$ is therefore

$$\cdot63183+\frac{\cdot08306}{3}+\frac{\cdot10635}{9}+\frac{\cdot07804}{27}+\frac{\cdot06489}{81}+\frac{\cdot05443}{243}+\frac{\cdot01437}{729}+\frac{\cdot01692}{2187}+\frac{\cdot01144}{6561}$$
$$+\frac{\cdot00367}{19683}+\frac{\cdot00104}{59049}+\frac{\cdot00015}{177147}+$$

The value of which to five places of decimals is ·67528.

The constant terms, therefore in f_1, f_2 up to f_5 when reduced to decimals, are in this case ·33333, ·48148, ·57110, ·64113, and ·65628 respectively. That is to say, out of a million surnames at starting, there have disappeared in the course of one, two, etc., up to five generations, 333333, 481480, 571100, 641130, and 675280 respectively.

The disappearances are much more rapid in the earlier than in the later generations. Three hundred thousand disappear in the first generation, one hundred and fifty thousand more in the second, and so on, while in passing from the fourth to the fifth, not more than thirty thousand surnames disappear.

All this time the male population remains constant. For it is evident that the male population of any generation is to be found by

multiplying that of the preceding generation, by $t_1 + 2t_2$, and this quantity is in the present case equal to one.

If axes Ox and Oy be drawn, and equal distances along Ox represent generations from starting, while two distances are marked along every ordinate, the one representing the total male population in any generation, and the other the number of remaining surnames in that generation, of the two curves passing through the extremities of these ordinates, the *population* curve will, in this case, be a straight line parallel to Ox, while the *surname* curve will intersect the population curve on the axis of y, will proceed always convex to the axis of x, and will have the positive part of that axis for an asymptote.

The case just discussed illustrates the use to be made of the general formula, as well as the labour of successive substitutions, when the expressions $f_1(x)$ does not follow some assigned law. The calculation may be infinitely simplified when such a law can be found; especially if that law be the expansion of a binomial, and only the extinctions are required.

For example, suppose that the terms of the expression $t_0 + t_1 x +$ &c. $+ t_q x_q$ are proportional to the terms of the expanded binomial $(a+bx)^q$ *i. e.*, suppose that $t_0 = \frac{a^q}{(a+b)^q}$ $t_1 = q\frac{a^{q-1}b}{(a+b)^q}$ and so on.

$$\text{Here } f_1(x) = \frac{(a+bx)^q}{(a+b)^q} \text{ and } {}_1m_0 = \frac{a^q}{(a+b)^q}$$

$$f_2(x) = \frac{1}{(a+b)^q}\left\{a + b\frac{(a+bx)^q}{(a+b)^q}\right\}^q$$

$${}_2m_0 = \frac{1}{(a+b)^q}\left\{a + b\,{}_1m_0\right\}^q$$

$$\text{Generally } {}_rm_0 = \frac{1}{(a+b)^q}\left\{a + b\,{}_{r-1}m_0\right\}^q = \frac{b^q}{(a+b)^q}\left\{\frac{a}{b} + {}_{r-1}m_0\right\}^q$$

If, therefore, we wish to find the number of extinctions in any generation, we have only to take the number in the preceding generation, add it to the constant fraction $\frac{a}{b}$, raise the sum to the power of q, and multiply by $\frac{b^q}{(a+b)^q}$

With the aid of a table of logarithms, all this may be effected for a great number of generations in a very few minutes. It is by no means unlikely that when the true statistical data t_0, t_1, etc., t_q are ascertained, values of a, b, and q may be found, which shall render the terms of the expansion $(a+bx)^q$ approximately proportionate to the terms of $f_1(x)$. If this can be done, we may *approximate* to the determination of the rapidity of extinction with very great ease, for any number of generations, however great.

For example, it does not seem very unlikely that the value of q might be 5, while t_0, $t_1 \ldots t_q$ might be ·237, ·396, ·264, ·088, ·014, ·001, or nearly, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{7}{24}$, $\frac{1}{23}$, $\frac{1}{138}$, and $\frac{1}{1000}$.

Should that be the case, we have $f_1(x)=\frac{(3+x)^5}{4^5}$ $_1m_0=\frac{3^5}{4^5}$

and generally $_rm_0=\frac{1}{4^5}\left\{3+{}_{r-1}m_0\right\}^5$

Thus we easily get for the number of extinctions in the first ten generations respectively

·237, ·346, ·410, ·450, ·477, ·496, ·510, ·520, ·527, ·533

We observe the same law noticed above in the case of $\frac{1+x+x^2}{3}$ viz., that while 237 names out of a thousand disappear in the first step, and an additional 109 names in the second step, there are only 27 disappearances in the fifth step, and only 6 disappearances in the tenth step.

If the curves of surnames and of population were drawn from this case, the former would resemble the corresponding curve in the case last mentioned, while the latter would be a curve whose distance from the axis of x increased indefinitely, inasmuch as the expression

$$t_1+2t_2+3t_3+4t_4+5t_5$$

is greater than one.

Whenever $f_1(x)$ can be represented by a binomial, as above suggested, we get the equation

$$_rm_0=\frac{1}{(a+b)^q}\left\{a+b_{r-1}m\right\}^q$$

whence it follows that as r increases indefinitely the value of $_rm_0$ approaches indefinitely to the value y where

$$y=\frac{1}{(a+b)}\left\{a+by\right\}$$

that is where $y=1$.

All the surnames, therefore, tend to extinction in an indefinite time, and this result might have been anticipated generally, for a surname once lost can never be recovered, and there is an additional chance of loss in every successive generation. This result must not be confounded with that of the extinction of the male population; for in every binomial case where q is greater than 2, we have t_1+2t_2+ &c. $+qt_q>1$, and, therefore an indefinite increase of male population.

The true interpretation is that each of the quantities, $_rm_1$, $_rm_2$, &c., tends to become zero, as r is indefinitely increased, but that it does not follow that the product of each by the infinitely large number N is also zero.

As, therefore, time proceeds indefinitely, the number of surnames extinguished becomes a number of the *same order of magnitude* as the total number at first starting in N, while the number of surnames

represented by one, two, three, etc., representatives is some infinitely smaller but finite number. When the finite numbers are multiplied by the corresponding number of representatives, sometimes infinite in number, and the products added together, the sum will generally exceed the original number N. In point of fact, just as in the cases calculated above to five generations, we had a continual, and indeed at first, a rapid extinction of surnames, combined in the one case with a stationary, and in the other case an increasing population, so is it when the number of generations is increased indefinitely. We have a continual extinction of surnames going on, combined with constancy, or increase of population, as the case may be, until at length the number of surnames remaining is absolutely insensible, as compared with the number at starting; but the total number of representatives of those remaining surnames is infinitely greater than the original number.

We are not in a position to assert from *actual calculation* that a corresponding result is true for every form of $f_1(x)$, but the reasonable inference is that such is the case, seeing that it holds whenever $f_1(x)$ may be compared with $\frac{(a+bx)_q}{(a+b)^q}$ whatever a, b, or q may be.

On the RUDE STONE MONUMENTS *of* CERTAIN NAGA TRIBES, *with some* REMARKS *on their* CUSTOMS, *etc.* By Major H. H. GODWIN-AUSTEN, F.R.G.S., F.Z.S., etc., Deputy Superintendent Topographical Survey of India. [With Plates xi and xii.]

ON visiting the Nágá Hills District last cold weather, 1872-73, I was very much surprised and interested to find that some of the tribes Anghámi and others erect upright cenotaphs, similar to those to be seen in the Khási Hills, and which I described when last in England in a paper read before this Institute in May 1871, and published in the Journal. The custom is here in full force, not, as is the case among the Khasis, undoubtedly fast dying out. The interest attached to this custom was not a little increased when I came on the first monoliths by my never having read of any notice of it in any work or report in which the Nágá tribes are mentioned. Colonel Butler, in his book, does not allude to this very remarkable custom, and Colonel Dalton is equally silent in his much later published work, "Descriptive Ethnology of Bengal." Not only are monolithic monuments common, but the Dolmen form is also to be seen in villages at the head of the Zúllo and Sijjo valleys. I first observed these stones on approaching the village of Kheruphima, set up on the roadside, often singly, in twos and threes, sometimes in sets of as many as