STATISTICAL PROPERTIES OF INVERSE GAUSSIAN DISTRIBUTIONS. I.

By M. C. K. Tweedie[1]

Virginia Agricultural Experiment Station, Virginia Polytechnic Institute

0. Summary. A report is presented on some statistical properties of the family of probability density functions

$$\exp\left[-\lambda(x-\mu)^2/2\mu^2x\right][\lambda/2\pi x^3]^{1/2}$$

for a variate x and parameters μ and λ, with x, μ, λ each confined to $(0, \infty)$. The expectation of x is μ, while λ is a measure of relative precision. The chief result is that the ml estimators of μ and λ have stochastically independent distributions, and are of a nature which permits of the construction of an analogue of the analysis of variance for nested classifications. The ml estimator of μ is the sample mean, and for a fixed sample size n its distribution is of the same family as x, with the same μ but with λ replaced by λn. The distribution of the ml estimator of the reciprocal of λ is of the chi-square type. The probability distribution of $1/x$, and the estimation of certain functions of the parameters in heterogeneous data, are also considered.

1. Introduction. The name "Inverse Gaussian" has been suggested [1] for the members of a certain family of continuous probability density functions in which the variate takes positive values only. The family is generated by varying the values of two real positive parameters, which may be any independent pair from α, λ, μ, ϕ, where $\frac{1}{2}\alpha^2 = \mu = \lambda/\phi$. The density function for the positive values of the variate may accordingly be written in the forms

$$\text{(1a)}\quad f_1(x; \alpha, \lambda) = \exp\{-\alpha\lambda x + \lambda(2\alpha)^{1/2} - \lambda/2x\}[\lambda/2\pi x^3]^{1/2}.$$

$$\text{(1b)}\quad f_2(x; \mu, \lambda) = \exp\{-\lambda(x-\mu)^2/2\mu^2 x\}[\lambda/2\pi x^3]^{1/2},$$

$$\text{(1c)}\quad f_3(x; \mu, \phi) = \exp\left\{-\frac{\phi x}{2\mu} + \phi - \frac{\mu\phi}{2x}\right\}[\mu\phi/2\pi x^3]^{1/2},$$

$$\text{(1d)}\quad f_4(x; \phi, \lambda) = \exp\left\{-\frac{\phi^2 x}{2\lambda} + \phi - \frac{\lambda}{2x}\right\}[\lambda/2\pi x^3]^{1/2}.$$

Each of these forms is convenient or suggestive for some purpose.

The relationships

$$\text{(2)}\qquad f_2(x; \mu, \lambda) = \mu^{-1} f_3(x/\mu; 1, \phi) = \lambda^{-1} f_4(x/\lambda; \phi, 1)$$

Received April 9, 1956; revised July 3, 1956.

[1] Present address: Statistical Laboratory, Department of Mathematics, The University, Manchester, England.

are useful in computing numerical values of the probability density. The cumulative probability function depends essentially on only two variables, which might be chosen to be x/μ and ϕ. The case $\mu = 1$ could therefore be adopted as a standard form. Curves of the density functions for $\lambda = \phi = \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16, 32$, with $\mu = 1$, are shown in Fig. 1. In some physical applications it is more natural to hold λ constant, and Fig. 2 shows the density curves for $\lambda = 1$ with $\mu = 4, 1$, and $\frac{1}{4}$, i.e., for $\phi = \frac{1}{4}$, 1, and 4 respectively.

Since it will be found useful to consider also some functions of the same algebraic form but with complex values for some of the parameters, it may be noted that the integrals of functions such as (1) over the interval $(0, \infty)$ can be shown to be unity, provided that the real parts of λ and of the mutually equal quantities $\alpha\lambda$ and $\frac{1}{2}\lambda\mu^{-2}$ are positive. For reference we reproduce an equation for a modified Bessel function of the second kind,

$$K_{\pm\nu}(z) = \tfrac{1}{2}(\tfrac{1}{2}z)^{\nu} \int_0^{\infty} \exp\left\{-t - \frac{z^2}{4t}\right\} \frac{dt}{t^{\nu+1}}, \tag{3}$$

given by Watson ([2], p. 183), under the condition that the real part of z^2 is positive, together with the result

$$K_{\pm 1/2}(z) = e^{-z}(\pi/2z)^{1/2}, \tag{4}$$

also given by Watson ([2], p. 80).

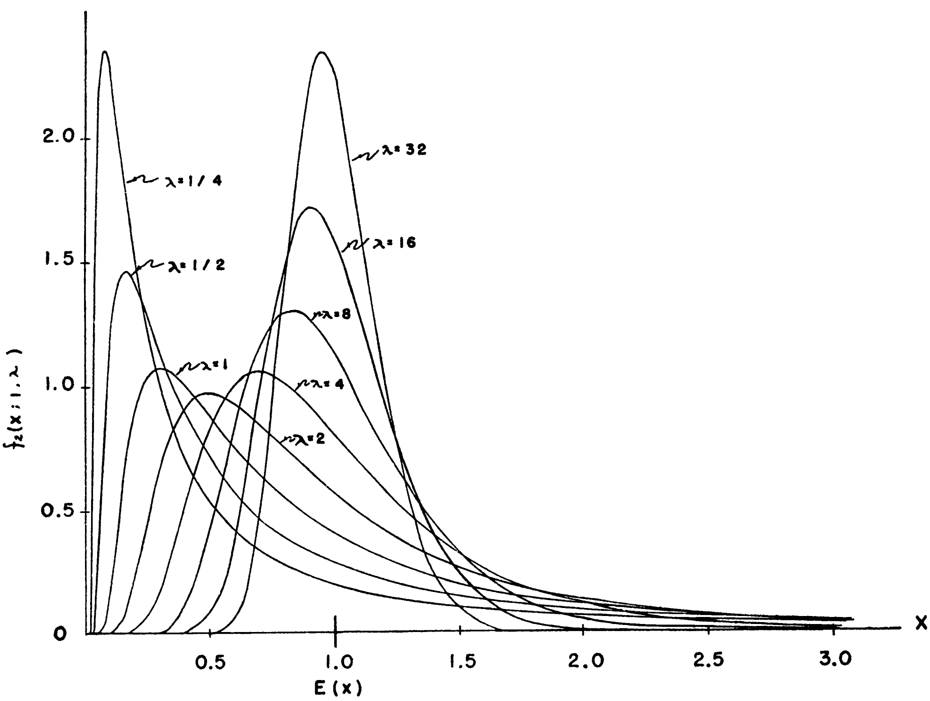

FIG. 1. Probability density curves for an Inverse Gaussian variate with $\mu = 1$ for 8 values of λ or ϕ.

The Inverse Gaussian family of distributions arises in a problem of Brownian motion (cf. [1], [3]), though then a further parameter appears in the physical formulation. The numerical value of this parameter can however normally be regarded as known, and it merely modifies the values of the parameters given in the expressions (1) above. Both λ and μ are of the same physical dimensions as the random variable x itself. A change of scale of x, such as may be due to a change in measuring unit or, approximately, to changes in temperature or some other factor, produces another member of the family, in which λ and μ have been multiplied by the same factor as x. The ratio ϕ is invariant under such a change.

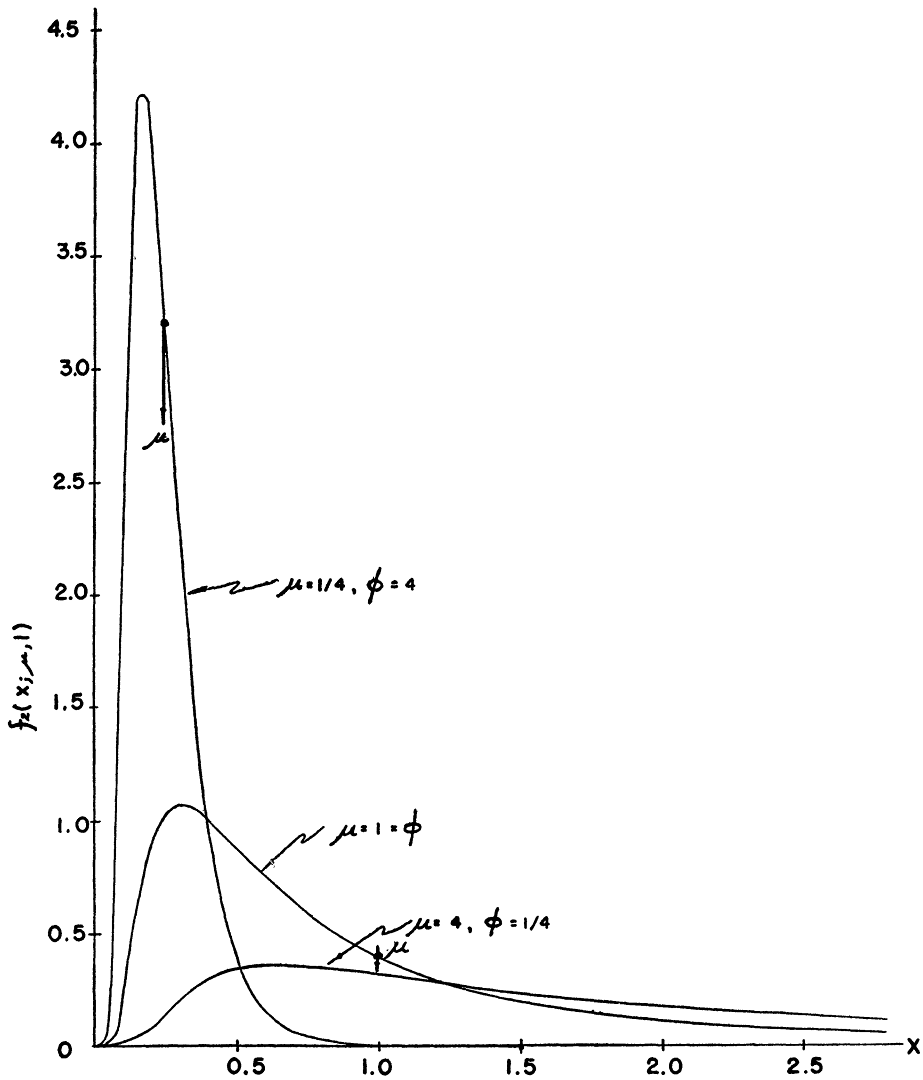

FIG. 2. Probability density curves for an Inverse Gaussian variate with $\lambda = 1$ for 3 values of μ or ϕ.

The same family appears also as a limiting form for the distribution of the final sample size in a special case of Wald's sequential likelihood ratio test [4]. Some properties of this family were studied in a degree thesis [5], where the Brownian motion problem was found to have an important part in the interpretation of some experimental work. The present paper establishes some of those exact properties in a more formal way, though using essentially the same methods as in the thesis. Some new results are included, and some further ones will be given in another paper. Not all these results are of quantitative importance in the original physical problem, and those which are not are presented here for their theoretical interest. The formulae (1) will be regarded as given, in that no derivations will be offered here. The uniqueness of certain Laplace transforms will be an important factor in some of the proofs. The form (1a) is of the kind adopted in an earlier published paper [6], in which similar methods were used.

2. Basic characteristics.

The shape of the distribution depends on ϕ only. The distribution is unimodal, with its mode at

$$x_{\text{mode}} = \mu\left\{\left(1 + \frac{9}{4\phi^2}\right)^{1/2} - \frac{3}{2\phi}\right\}. \tag{5}$$

The ratio x_{mode}/μ converges to 1 when ϕ is increased to infinity; while the ratio x_{mode}/λ converges to $\frac{1}{3}$ when ϕ decreases to zero. The density at the mode is least when $\phi = 2$, if μ is fixed. The mode then occurs at $x = \frac{1}{2}\mu$ and the density there is $[8/\pi\mu^2 e]^{1/2} = 0.96788\mu^{-1}$.

It is convenient to introduce the logarithm of the Laplace transform $E(e^{-tx})$ of the probability density of the variate, which is in a sense a cumulant-generating function (cgf). We denote the relevant function operator by L, with the variate symbol as a subscript and the other variables in parentheses. Thus, from form (1a),

$$L_x(t; \mu, \lambda) = \ln e^{\lambda(2\alpha)^{1/2}} \int_0^\infty e^{-(\alpha+t/\lambda)\lambda x - \lambda/2x}\, dx[\lambda/2\pi x^3]^{1/2} \tag{6}$$

$$= \lambda(2\alpha)^{1/2} - \lambda 2^{1/2}\left(\alpha + \frac{t}{\lambda}\right)^{1/2} + \ln \int_0^\infty f_1\left(x; \alpha + \frac{t}{\lambda}, \lambda\right) dx. \tag{7}$$

If t is imaginary, or, if real or complex, if its real part exceeds $-\alpha\lambda$, the integral in (7) is unity. Hence

$$L_x(t; \mu, \lambda) = \lambda\left\{(2\alpha)^{1/2} - 2^{1/2}\left(\alpha + \frac{t}{\lambda}\right)^{1/2}\right\} \tag{8a}$$

$$= \frac{\lambda}{\mu}\left\{1 - \left(1 + \frac{2\mu^2 t}{\lambda}\right)^{1/2}\right\} \tag{8b}$$

$$= \phi\left\{1 - \left(1 + \frac{2\mu t}{\phi}\right)^{1/2}\right\} \tag{8c}$$

$$= \phi\left\{1 - \left(1 + \frac{2\lambda t}{\phi^2}\right)^{1/2}\right\}. \tag{8d}$$

This cgf. is unique to the density function (1).

The cumulants can be obtained from the power series expansion of $L_x(t; \mu, \lambda)$. They are:

$$\kappa_1 = \mu = \lambda\phi^{-1}, \qquad \kappa_2 = \mu^3\lambda^{-1} = \lambda^2\phi^{-3},$$
$$\kappa_3 = 3\mu^5\lambda^{-2} = 3\lambda^3\phi^{-5}, \qquad \kappa_4 = 15\mu^7\lambda^{-3} = 15\lambda^4\phi^{-7},$$

and, in general, when $r \geqq 2$,

$$\begin{aligned}\kappa_r &= 1\cdot3\cdot5 \cdots (2r-3)\mu^{2r-1}\lambda^{1-r}\\ &= \lambda^r(2r-3)!/\phi^{2r-1}2^{r-2}(r-2)!.\end{aligned} \tag{9}$$

Thus μ is the population mean and is primarily a measure of location, while λ is an inverse measure of relative dispersion, being the ratio of κ_1^3 to κ_2, or

$$\frac{1}{\lambda} = \frac{\kappa_2}{\kappa_1^3}. \tag{10}$$

Also, $\phi = \kappa_1^2/\kappa_2$. The Fisherian shape coefficients, or standardized cumulants, are

$$\begin{aligned}\gamma_1 &= \kappa_3\kappa_2^{-3/2} = 3\phi^{-1/2}, \qquad \gamma_2 = \kappa_4\kappa_2^{-2} = 15\phi^{-1},\\ \gamma_r &= \kappa_{r+2}\kappa_2^{-r/2-1} = 1\cdot3\cdot5 \cdots (2r+1)\phi^{-r/2}.\end{aligned} \tag{11}$$

The fractional coefficient of variation is $\gamma_{-1}^{-1} = \kappa_1^{-1}\kappa_2^{1/2} = \phi^{-1/2}$, so that $\gamma_1 = 3\gamma_{-1}^{-1}$, $\gamma_2 = 15\gamma_{-1}^{-2}$, and so on. Evidently the distribution becomes more and more nearly normal when ϕ is increased. This parameter ϕ might be called the normality parameter or the shape parameter.

In the probability density curves shown in Fig. 1, γ_1 ranges from 6 down to 0.53, and γ_2 ranges from 60 down to 0.47. The approach to normality in the neighbourhood of $x = \mu$ is evident from these curves. However, some important aspects of the distributions, such as the standardized cumulants, depend primarily on the behaviour of the functions at very large values of the variate, whereas the diagrams are necessarily bounded.

The positive integral moments about zero are obtainable either from (9) or by direct integration, using (3) and a further result given by Watson ([2], Eq. (12), p. 80). They are:

$$\begin{aligned}\mu_1' &= \mu, \qquad \mu_2' = \mu^2 + \mu^3\lambda^{-1},\\ \mu_3' &= \mu^3 + 3\mu^4\lambda^{-1} + 3\mu^5\lambda^{-2},\\ \mu_4' &= \mu^4 + 6\mu^5\lambda^{-1} + 15\mu^6\lambda^{-2} + 15\mu^7\lambda^{-3},\\ \mu_r' &= \mu^r K_{r-1/2}(\phi)K_{1/2}(\phi) = \mu^r \sum_{s=0}^{r-1} \frac{(r-1+s)!}{s!(r-1-s)!(2\phi)^s}.\end{aligned} \tag{12}$$

The negative integral moments are given in (33).

It follows from the form of (8) that the distribution of the arithmetic mean of a fixed number n of independent values from (1) is a member of the same family, with the same α and μ, but with λ replaced by λn and ϕ replaced by ϕn. More generally, suppose that we have a set of populations in which μ_i and λ_i are the values of the parameters in the i-th population, and that, although the values of these parameters are unknown, the values of $a_i = C\mu_i^{-2}\lambda_i$ are known, C being a constant whose value is not necessarily known. The distribution of the linear function $\sum_{i=1}^{n} (a_i x_i)$ is then of the same form as (1), with $\mu = C\sum_{i=1}^{n} \phi_i$, $\lambda = C(\sum_{i=1}^{n} \phi_i)^2$, $\phi = \sum_{i=1}^{n} \phi_i$. Because of this additive property of the normality parameter, the linear function will have a more nearly normal distribution than any of its components.

3. Estimation of parameters. Suppose that x_i is an observation on a distribution of the form (1b), with parameter values μ and λ_i, where $\lambda_i = \lambda_0 w_i$ for $i = 1$ to N, w_i being positive and known, but neither of the common values of μ and λ_0 being known. For example, x_i might be the arithmetic mean of w_i values from a distribution with parameter values μ and λ_0. With these N pairs of values of x_i and w_i as data, the estimates of μ and λ_0 which jointly maximize the likelihood function are given by

$$\hat{\mu} = x_. = \sum_{i=1}^{N} (w_i x_i) \Big/ \sum_{i=1}^{N} (w_i), \tag{13}$$

$$\frac{1}{\hat{\lambda}_0} = \frac{1}{N} \sum_{i=1}^{N} w_i \left(\frac{1}{x_i} - \frac{1}{x_.}\right). \tag{14}$$

These estimates can never be negative so long as the observations are necessarily non-negative. For (14) this follows from the concavity of the function x^{-1}. With every w_i equal to unity, these estimates were given by Schrödinger [3], who called them "wahrscheinlichste."

The Inverse Gaussian family is one for which the weighted sample mean $x_. = \hat{\mu}$ (13) is a sufficient statistic (in Fisher's sense) for estimating the common population mean μ. Further, the cumulant-generating function of $\hat{\mu}$, with fixed values of μ, λ_0, N and the weights $w_1, \cdots w_N$, differs from (8) only in that λ becomes $\lambda_0 \sum_{i=1}^{N} w_i$. (To see this, take $C = \mu^2/\lambda_0 \sum_{i=1}^{N} w_i$ in the result at the end of Section 2.) The probability density function of $\hat{\mu}$ therefore is

$$f_2(\hat{\mu}; \mu, \lambda_0 \textstyle\sum_{i=1}^{N} w_i), \qquad 0 < \hat{\mu} < \infty. \tag{15}$$

In the terminology of a previous paper [6], the family (1) is a Laplacian one with α as primary parameter and λ as secondary parameter. Hence $\hat{\mu}$ has a Laplacian form of probability density function. This enables the conditional moments and cumulants and other properties of other statistics, with a fixed value of $\hat{\mu}$, to be found by using the uniqueness of the Laplace transforms which appear in their mathematical formulations. A number of exact results have been found for the Inverse Gaussian distributions in this way, and we shall now proceed to develop one of the more surprising of them.

4. Distribution of the ml estimator of the secondary parameter. With the same data and the same fixed quantities as were introduced in Section 3 in discussing the distribution of the maximum likelihood (ml) estimator of μ, the Laplace transform of the probability density function of $1/\hat{\lambda}_0$ is

$$E(e^{-t/\hat{\lambda}_0}) = \underset{\text{all } X_i>0}{\int \cdots \int} e^{-t/\hat{\lambda}_0} \prod_{i=1}^{N} f_2(X_i;\, \mu,\, \lambda_0 w_i)\, dX_i. \tag{16}$$

This certainly exists when the real part of t is not negative. On substituting from (1) and (15) and writing $\sum_{i=1}^{N} w_i = W$ for brevity, we get

$$\begin{aligned} E(e^{-t/\hat{\lambda}_0} \mid \mu, \lambda_0, w_1, \cdots w_N, N) = \int_{\hat{\mu}=0}^{\hat{\mu}=\infty} f_2(\hat{\mu}; \mu, \lambda_0 W) \underset{\hat{\mu}\text{ constant}}{\int \cdots \int} e^{-(t+\lambda_0 N/2)/\hat{\lambda}_0} \\ \cdot \frac{\hat{\mu}^{3/2}}{W^{1/2}} \left(\frac{\lambda_0}{2\pi}\right)^{(N-1)/2} \prod_{i=1}^{N} \frac{w_i^{1/2}\, dX_i}{X_i^{3/2}}. \end{aligned} \tag{17}$$

The multiple integral in the final integrand on the right of (17) does not contain μ or α. From the Laplacian form of $f_2(\hat{\mu};\, \mu,\, \lambda_0 W)$ and the uniqueness of the Laplace transforms to which it gives rise, it follows (cf. [6]) that the partial derivative, with respect to $\hat{\mu}$, of this multiple integral is equal to the Laplace transform of the conditional density of $1/\hat{\lambda}_0$. This statement may be justified by reference either to Lerch's theorem ([7], p. 52; [8], p. 61) or to an equally applicable set of theorems (cf. [9], p. 38). The proof can legitimately involve α taking complex values with positive real parts. Therefore

$$\begin{aligned} \frac{\partial}{\partial \hat{\mu}} \underset{\hat{\mu}\text{ constant}}{\int \cdots \int} e^{-(t+\lambda_0 N/2)/\hat{\lambda}_0} \hat{\mu}^{3/2} W^{-1/2} (\lambda_0/2\pi)^{(N-1)/2} \prod_{i=1}^{N} (X_i^{-3/2} w_i^{1/2}\, dX_i) \\ = E(e^{-t/\hat{\lambda}_0} \mid \hat{\mu}; \lambda_0, w_1, \cdots w_N, N), \end{aligned} \tag{18}$$

which is the conditional moment-generating function of $1/\hat{\lambda}_0$ with $\hat{\mu}$ fixed. To evaluate this integral, first take $t = 0$, which gives

$$\frac{\partial}{\partial \hat{\mu}} \underset{\hat{\mu}\text{ constant}}{\int \cdots \int} e^{-\lambda_0 N/2/\hat{\lambda}_0} \hat{\mu}^{3/2} \prod_{i=1}^{N} (X_i^{-3/2}\, dX_i) = \frac{W^{1/2}}{\prod_{i=1}^{N} (w_i^{1/2})} \left(\frac{2\pi}{\lambda_0}\right)^{(N-1)/2} \tag{19}$$

By substituting $\lambda_0 + 2t/N$ for λ_0 on both sides of (19), the left-hand side of (18) is found almost immediately to be

$$E(e^{-t/\hat{\lambda}_0} \mid \hat{\mu}; \lambda_0, w_1, \cdots w_N, N) = \left(1 + \frac{2t}{\lambda_0 N}\right)^{(N-1)/2} \tag{20}$$

This is a Laplace transform of a density function of the chi-square type, with N-1 degrees of freedom. In fact, it shows that

$$\frac{1}{\hat{\lambda}_0} = \frac{\chi^2_{(N-1\text{ d.f.})}}{\lambda_0 N} \tag{21}$$

and thus that

$$\sum_{i=1}^{N} w_i \left(\frac{1}{x_i} - \frac{1}{x_.}\right) = \frac{\chi^2_{(N-1\text{ d.f.})}}{\lambda_0}. \tag{22}$$

This result can be used to obtain confidence intervals for λ_0. By substituting for the probability elements of $\hat{\mu}$ and $\hat{\lambda}_0$ in the joint probability element of the N observations, it can be shown also that $\hat{\mu}$ and $\hat{\lambda}_0$ are jointly sufficient estimators of μ and λ_0. Further,

$$\frac{1}{N-1} \sum_{i=1}^{N} w_i \left(\frac{1}{x_i} - \frac{1}{x_.}\right) \tag{23}$$

is an unbiased estimator of $1/\lambda_0$. Its distribution is of exactly the same type as that of the usual unbiased quadratic estimator of the variance of a Gaussian distribution, although it cannot be expressed precisely as a sum of squares of Gaussian variates with zero means. The conditional distribution of (23) is necessarily independent of μ, because of the sufficiency of $\hat{\mu}$, but it is also independent of $\hat{\mu}$, thus affording the possibility of an analogue of the analysis of variance, using the existing tables of χ^2 and F for significance tests and so forth. In the Brownian motion problem μ and λ are concerned with rather different physical properties of the experimental system, which do however occur together in some physical formulae. It is therefore convenient that estimators have been found which are both independently distributed and jointly sufficient.

The statistic (14) appears also in the likelihood ratio test of the hypothesis that the population means are equal against the alternative hypothesis that they may have any values independently of one another, the values for the means not being specified, while the value of the secondary parameter is supposed known. The logarithm of the ratio of the maximum likelihoods under these two hypotheses is $-\lambda_0 N/2\hat{\lambda}_0$, so that the result (21) is a case where the well-known approximate general result ([9], p. 151) holds exactly. Moreover, the statistic $\hat{\lambda}_0$ depends essentially on the difference between the ml estimator of the reciprocal of the hypothetically common value for μ and the ml estimator of the weighted mean of the reciprocals of the means under the assumption of their complete independence. Both these considerations indicate that $\hat{\lambda}_0$ will tend to be increased by real differences between the population means, and that it therefore measures the combined effect of the dispersion in homogeneous samples and the heterogeneity of the means.

5. An analogue of the analysis of variance for nested classifications. The algebraic aspect of the analysis of variance for nested classifications may be generalized, for two classifications (which is a sufficiently general case), to

$$\begin{aligned} &\textstyle\sum_{i=1}^{N}\sum_{j=1}^{n_i} \{\psi(x_{ij}) - \psi(x_{..})\} \\ &\quad = \textstyle\sum_{i=1}^{N}\sum_{j=1}^{n_i} \{\psi(x_{ij}) - \psi(x_{i.})\} + \sum_{i=1}^{N} n_i\{\psi(x_{i.}) - \psi(x_{..})\}. \end{aligned} \tag{24}$$

Here x_{ij} is the observed value of the variate in the j-th subclass of the i-th major class, and n_i is the number of subclasses in this major class, while there are N major classes. The values of $x_{i.}$ and $x_{..}$ are means of some kind of the values of x_{ij}, and $\psi(x)$ is some suitable function of x. The sums might be regarded as depending on the differences between different kinds of means, in that they could be rewritten as

$$\left\{\frac{1}{n_.}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\psi(x_{ij}) - \psi(x_{..})\right\}$$

(25)

$$= \frac{1}{n_.}\sum_{i=1}^{N} n_i \left\{\frac{\sum_{j=1}^{n_i}\psi(x_{ij})}{n_i} - \psi(x_{i.})\right\} + \left\{\frac{1}{n_.}\sum_{i=1}^{N}[n_i\psi(x_{i.})] - \psi(x_{..})\right\},$$

where $n_. = \sum_{i=1}^{N} n_i$. That is to say, if we temporarily use M to stand for the operation of taking the relevant mean involved in $x_{i.}$ and $x_{..}$, and A to stand for the operation of taking the weighted arithmetic mean, the identity (25) can be written

(26) $$A_iA_j\psi - \psi M_iM_j = A_i(A_j\psi - \psi M_j) + (A_i\psi - \psi M_i)M_j,$$

operating on x_{ij}. Certain restrictions on ψ and M are necessary to ensure that these differences, which are essentially measures of dispersion, shall never change sign. In the analysis of variance, the means entailed by M are arithmetical, while $\psi(x) = x^2$. If the variates have independent Gaussian distributions with both means and variances equal within any major group, the two major sums on the right side of (24) have independent distributions. This independence does not generally occur in other circumstances, but it is available to some extent with the Inverse Gaussian distribution. For this, according to Section 4, we again take the means M to be arithmetical, but take $\psi(x) = x^{-1}$.

From the results obtained in Section 4, we see that the statistic

(27) $$\sum_{j=1}^{n_i}\left(\frac{1}{x_{ij}} - \frac{1}{x_{i.}}\right) = \frac{n_i}{\hat{\lambda}_i}$$

is distributed as χ^2/λ_i with $n_i - 1$ degrees of freedom and independently of $x_{i\cdot}$, $= \sum_{j=1}^{n_i}(x_{ij})/n_i$. Hence

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{1}{x_{ij}} - \frac{1}{x_{i.}}\right)$$

is distributed as χ^2/λ with $n_. - N$ degrees of freedom and independently of the values of $x_{i\cdot}$. (This distribution would remain true even if the expectations $E(x_{i.})$ varied with i.) In particular, if $x_{..} = \sum_{i=1}^{N}(n_ix_{i.})/n_i$, that double sum is distributed independently of

$$\sum_{i=1}^{N} n_i\left(\frac{1}{x_{i.}} - \frac{1}{x_{..}}\right),$$

which is itself distributed as χ^2/λ with $(N - 1)$ degrees of freedom.

The algebraic identity (24) thus becomes

$$(28) \qquad \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{1}{x_{ij}} - \frac{1}{x_{..}}\right) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{1}{x_{ij}} - \frac{1}{x_{i.}}\right) + \sum_{i=1}^{N} n_i\left(\frac{1}{x_{i.}} - \frac{1}{x_{..}}\right).$$

If all the observations come independently from the same Inverse Gaussian distribution, the three major sums in (28) are each distributed as χ^2/λ, the chi-squares having respectively $n. - 1$, $n. - N$, and $N - 1$ degrees of freedom. Thus $1/\lambda$ can be estimated by dividing any of the three sums in (28) by the corresponding number of degrees of freedom. The two sums on the right of (28) have independent distributions and therefore their ratio is distributed as

$$(N - 1)F/(n. - N),$$

where F has $N - 1$ and $n. - N$ degrees of freedom. Hence the analogy with the analysis of variance is very close. For example, a significance test of the differences between the N values of $x_{i.}$ may be made by using the first major sum on the right of (28) as the analogue of the sum of squares for error. Some illustrations of the use of these formulae will be published separately for some electrophoretic data on individual colloid particles and for some purely empirical trials on more general data.

This "analysis of reciprocals" by (28) is invariant under changes of scale of the observations, but not under more general linear transformations, whereas the analysis of variance is thus invariant. It should also be noted that the obvious parallel with the algebraic identity for the main effects and interactions in the analysis of variance for crossed classifications does not give independent components. An interaction term, such as

$$\sum_{i=1}^{N}\sum_{j=1}^{n}\left(\frac{1}{x_{ij}} - \frac{1}{x_{i.}} - \frac{1}{x_{.j}} + \frac{1}{x_{..}}\right)$$

in a commonly used notation, does not have a distribution of the chi-square type, since it has a finite probability of taking a negative value, and therefore this analogue of the analysis of variance is restricted to nested classifications.

6. Distribution of the reciprocal of an Inverse Gaussian variate. For some purposes it is convenient to work with the reciprocal of the Inverse Gaussian variate x, which will be denoted by y. For example, the analysis discussed in Section 5 can be expressed simply in terms of this variable. The weighted arithmetic means $x_{i.}$, $x_{..}$ of the values of x_{ij} are replaced by their reciprocals, which are the weighted harmonic means $\tilde{y}_{i.}$, $\tilde{y}_{..}$ of the values of y_{ij}. The analysis of the values of y_{ij}, corresponding to the algebraic identity (24) or (26), with $\psi(y) = y$ and M the harmonic mean, thus becomes

$$(29) \qquad \sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \tilde{y}_{..}) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \tilde{y}_{i.}) + \sum_{i=1}^{N} n_i(\tilde{y}_{i.} - \tilde{y}_{..}).$$

These sums of course have the same chi-square types of distribution as the expressions in terms of x_{ij}, $x_{i.}$, and $x_{..}$, to which they are equal. However, this analysis (29) is sufficiently easy to compute to be considered as a further practical analogue of the analysis of variance for certain purposes.

Some of the useful properties of the variate y follow in an obvious way from those of x, hardly justifying giving any special consideration to the family of distributions of y. However, the latter has some interesting features and a short account is therefore in order. Some of the results will be expressed in terms of x, since that variate is the primary object of this study.

The probability density function of y may be written

$$\exp\left\{-\frac{\lambda y}{2}+\frac{\lambda}{\mu}-\frac{\lambda}{2\mu^2 y}\right\}[\lambda/2\pi y]^{1/2}, \qquad 0 < y < \infty \tag{30}$$

$$= \mu y f_2(y; \mu^{-1}, \lambda\mu^{-2}) = \mu^2 y f_3(\mu y; 1, \phi) = \mu^2 y f_4(\mu y; \phi, \phi). \tag{31}$$

The mode is at

$$y_{\text{mode}} = \frac{1}{2\lambda}[-1 + (1 + 4\phi^2)^{1/2}] = \frac{1}{\mu}\left\{\left(1 + \frac{1}{4\phi^2}\right)^{1/2} - \frac{1}{2\phi}\right\}.$$

The probability density at the mode is

$$\mu\{[1 + (1 + 4\phi^2)^{1/2}]/4\pi\}^{1/2} \exp\{\phi - (\phi^2 + \tfrac{1}{4})^{1/2}\},$$

which approaches $\mu/(2\pi e)^{1/2} = 0.241971\mu$ as its limit when ϕ or λ decreases to zero with a fixed value of μ.

Fig. 3 shows some examples of (30) plotted for $\lambda = 0, \frac{1}{4}, 1, 4, 16, 32$, with $\mu = 1$, for $0 < y < 3$, corresponding to $\frac{1}{3} < x < \infty$. The difference between Figs. 1 and 3 for small values of λ is rather striking.

Fig. 4 shows density curves from (30) in a form comparable with Fig. 2, having $\lambda = 1$ with $\mu = \frac{1}{4}, 1, 4$ for $0 < y < 7$. Thus the harmonic sample mean is a sufficient statistic for discriminating between the distributions of the family to which the curves in Fig. 4 belong, while the arithmetic mean is the corresponding statistic for Fig. 2 (cf. Section 3). Some consequences of using the arithmetic mean of observed values of y to estimate $1/\mu$ instead of using the harmonic mean are discussed in Section 7.

The moments about zero of y, which are the moments about zero of negative order for the Inverse Gaussian variate x, may be found by direct integration, using (3), or from other results. They may be found from those of x of positive order by using the relationship

$$E[(x/\mu)^{-r}] = E[(x/\mu)^{r+1}],$$

from which, in a notation applying to the variate x,

$$\mu'_{-r} = \mu'_{r+1}/\mu^{2r+1}. \tag{32}$$

Thus the moments of all positive and negative orders exist for an Inverse Gaussian variate (and for its reciprocal), in contrast to the situation with some other

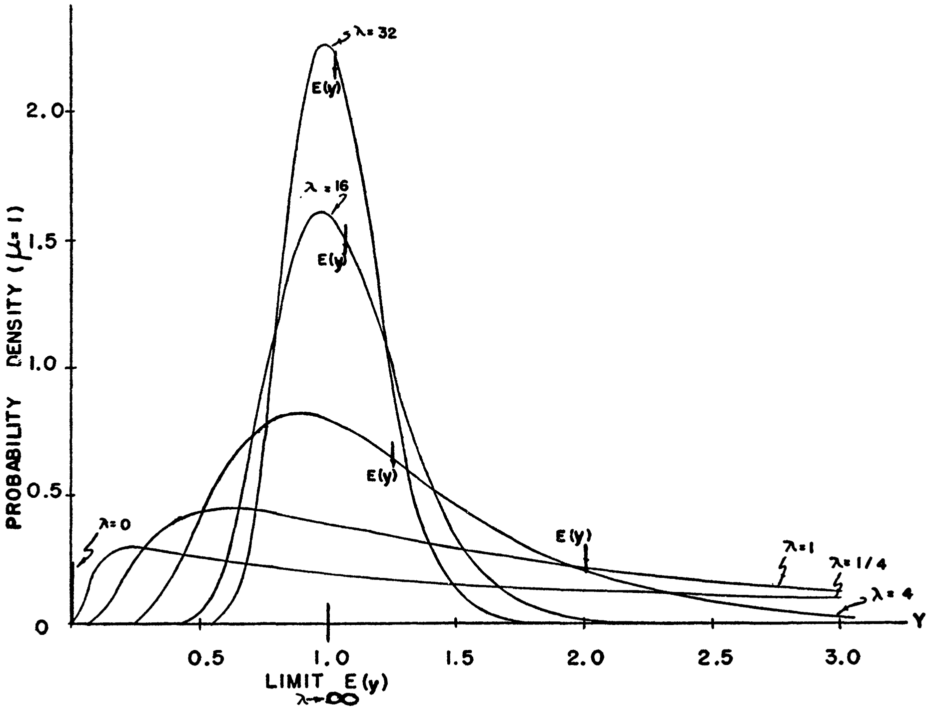

FIG. 3. Probability density curves for the reciprocal of an Inverse Gaussian variate with $\mu = 1$ for 6 values of λ or ϕ.

superficially similar distributions, such as the chi-square type. For reference, we have

$$\mu'_{-1} = \mu^{-1} + \lambda^{-1},$$

$$\mu'_{-2} = \mu^{-2} + 3\mu^{-1}\lambda^{-1} + 3\lambda^{-2},$$

$$\mu'_{-3} = \mu^{-3} + 6\mu^{-2}\lambda^{-1} + 15\mu^{-1}\lambda^{-2} + 15\lambda^{-3},$$

(33) $$\mu'_{-4} = \mu^{-4} + 10\mu^{-3}\lambda^{-1} + 45\mu^{-2}\lambda^{-2} + 105\mu^{-1}\lambda^{-3} + 105\lambda^{-4},$$

$$\mu'_{-r} = (2\lambda)^{-r} \sum_{s=0}^{r} \frac{(2r-s)!}{s!(r-s)!} (2\phi)^s$$

$$= \mu^{-r} \sum_{s=0}^{r} \frac{(r+s)!}{s!(r-s)!} (2\phi)^{-s}.$$

The family (30) is of the Laplacian form as regards the variate y. Thus its cgf. can be found by a process of substituting alternative values for the parameters in the integral of (30), in a similar way to that used for deriving the corresponding function (8) for x. The result is that the logarithm of the Laplace transform of the density function of y is

(34) $$L_y(t; \phi, \lambda) = \phi(1 - (1 + 2t\lambda^{-1})^{1/2}) - \tfrac{1}{2}\ln(1 + 2t\lambda^{-1}).$$

It is curious that the form of this function (34) shows that the distribution of y is the same as that of the convolution or composition of an Inverse Gaussian distribution (with the same value of ϕ but with μ replaced by $1/\mu$) with an independent distribution of χ^2/λ, this χ^2 having one degree of freedom.

The first two cumulants of y are

$$\kappa_1(y) = \frac{1}{\mu} + \frac{1}{\lambda}, \qquad \kappa_2(y) = \frac{1}{\lambda\mu} + \frac{2}{\lambda^2}. \tag{35}$$

Thus λ^{-1} is the bias in using y (or x^{-1}) as an estimator of μ^{-1}. The variate y itself may be the harmonic sample mean of values of a similar variate, or the reciprocal of the weighted arithmetic sample mean of values of Inverse Gaussian variates as in Section 3. The mean squared error in using y as an estimator of μ^{-1} is

$$E[(y - \mu^{-1})^2] = (\phi + 3)\lambda^{-2}. \tag{36}$$

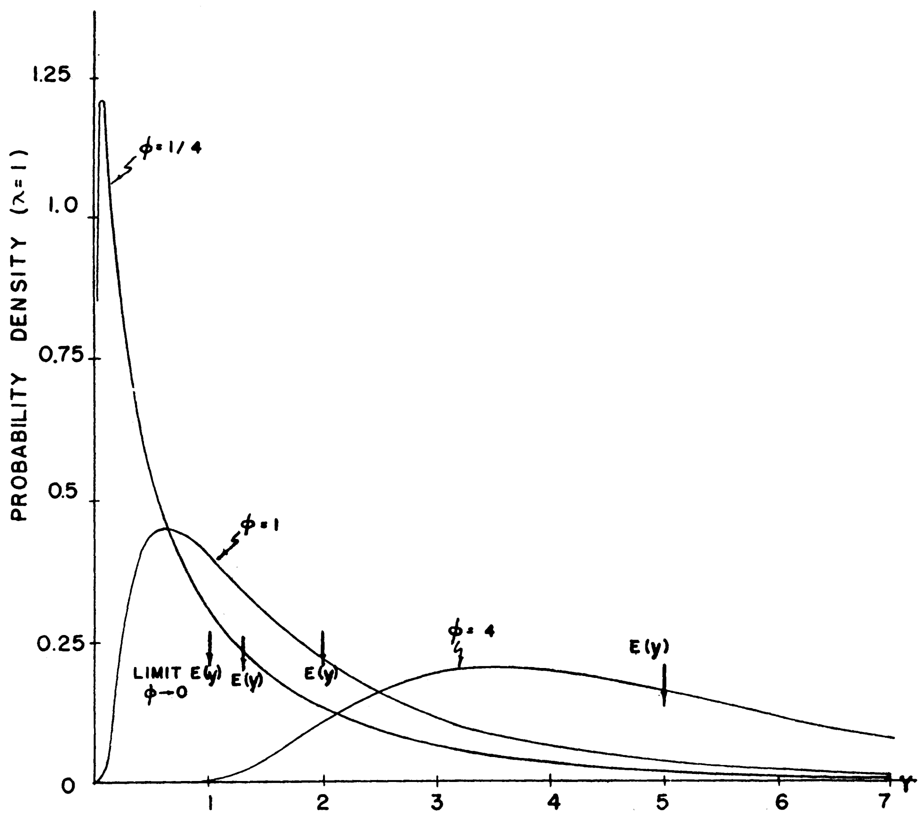

FIG. 4. Probability density curves for the reciprocal of an Inverse Gaussian variate with $\lambda = 1$ for 3 values of μ or ϕ.

The first two Fisherian shape coefficients of y are

$$\gamma_1(y) = (3\phi + 8)(\phi + 2)^{-3/2}, \tag{37}$$

$$\gamma_2(y) = 3(5\phi + 16)(\phi + 2)^{-2}. \tag{38}$$

The fractional coefficient of variation is

$$\gamma_{-1}(y)^{-1} = (\phi + 1)^{-1}(\phi + 2)^{1/2} \tag{39}$$

The values of γ_{-1}^{-1}, γ_1, and γ_2 are smaller than the values of the corresponding characteristics of x. When ϕ is increased, both these and the shape coefficients of higher order approach zero as their limit, so that the distribution of y then approaches normality.

7. Estimation of the arithmetic mean reciprocal of expectations of Inverse Gaussian variates. In the physical experiments which led to this research, it was desired to estimate the arithmetical mean reciprocal of the population means of four distributions which might reasonably be treated as Inverse Gaussian. The four means were not individually of special interest, their inequality being due to an artifact of the measuring technique. The secondary parameter λ could be considered to be constant in any one experiment. This estimation problem may be discussed in the following more general terms:

Suppose that we have N populations of the kind (30), the i-th having parameter values μ_i, $\lambda_i = \lambda_0 w_i$, and one observation y_i from each population. Write $\bar{y} = \sum_{i=1}^{N} (w_i y_i)/\sum_{i=1}^{N} (w_i)$. It follows from the form of (34) that the distribution of $\bar{y}$ is the same as that of the simple arithmetic mean of N values of y taken from one distribution whose parameter values are

$$\mu^* = \sum_{i=1}^{N} (w_i)/\sum_{i=1}^{N} (w_i/\mu_i), \qquad \lambda^* = \lambda_0 \sum_{i=1}^{N} (w_i)/N.$$

We may also write $\phi^* = \lambda^*/\mu^*$. The distribution of $\bar{y}$ is the convolution of an Inverse Gaussian distribution, whose parameters in the form (1b) have the values $1/\mu^*$, $\lambda^* N$, with an independent distribution of $\chi^2/\lambda^* N$, this chi-square having N degrees of freedom. Because this belongs to a different Laplacian family it will not be studied in detail here. The results needed at present can be obtained from those already found.

From (35),

$$E(\bar{y}) = (1 + 1/\phi^*)/\mu^*, \tag{40}$$

and

$$E[(\bar{y} - 1/\mu^*)^2] = (\phi^* + 2 + N)/\lambda^{*2} N. \tag{41}$$

If the N values of μ_i were all equal, the harmonic sample mean

$$\tilde{y} = \sum_{i=1}^{N} (w_i)/\sum_{i=1}^{N} (w_i/y_i)$$

would be a more precise estimator of the common value. The formula (36) would then give

$$(42) \qquad E[(\tilde{y} - 1/\mu)^2] = (\phi^* + 3/N)/\lambda^{*2}N,$$

which is less than (41) except in the trivial case of $N = 1$. The efficiency of $\tilde{y}$, in these circumstances, can be measured by the ratio of the mean squared errors (42) and (41), or by the ratio, to N, of the modified value for N which needs to be substituted (without changing λ^* and ϕ^*) in (42) to give a mean squared error for $\tilde{y}$ equal to (41). The former measure of efficiency is easier to calculate, but is slightly less than the latter. However, the difference is less than one per cent if $\phi^* > 18.7$.

Reverting to the estimation of μ^* when the values of μ_i may be unequal, we may attempt to improve the estimator $\bar{y}$ by adjusting it for the bias. When the N values of y_i are the harmonic sample means of the reciprocals of Inverse Gaussian variates, or the reciprocals of arithmetic sample means of Inverse Gaussian variates, separate estimates of the values of λ_i can be obtained from the variation exhibited within the samples, by using the appropriate form of (27). If the i-th sample contains n_i observations,

$$\hat{\lambda}_i^{-1} = \chi^2_{(n_i - 1 \text{ d.f.})}/\lambda_0 n_i w_i ,$$

this distribution being independent of y_i. On weighting these estimators suitably and on writing $\sum_{i=1}^N (n_i - 1) = D$, we have, as an unbiased estimator of λ_0^{-1} of minimum variance,

$$\sum_{i=1}^N (\hat{\lambda}_i^{-1} n_i w_i)/D = \chi^2_{(D \text{ d.f.})}/\lambda_0 D.$$

Thus, for an unbiased estimator of $1/\mu^*$, we get

$$(43) \qquad y' = \sum_{i=1}^N w_i(y_i - n_i N/\hat{\lambda}_i D)/\sum_{i=1}^N (w_i).$$

Then

$$(44) \qquad E[(y' - 1/\mu^*)^2] = (\phi^* + 2 + 2N/D)/\lambda^{*2}N.$$

This mean squared error will always be less than that (41) of $\bar{y}$ if $D > 2$, which will be true of most experiments. However, unless ϕ^* is close to or less than unity, which seems unlikely to occur, the statistical superiority of y' over $\bar{y}$ is of very minor importance.

8. Acknowledgments. Much of this work was done under a Senior Scholarship at the University of Reading, England, followed by financial assistance from the British Empire Cancer Campaign. The recent developments were made under the sponsorship of the National Science Foundation. The figures were drawn by Mrs. D. Hamilton from tables of values which were computed by Mr. G. Zorbalas.

REFERENCES

[1] M. C. K. TWEEDIE, "Inverse statistical variates," *Nature* (London), Vol. 155 (1945), p. 453.

[2] G. N. WATSON, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, 1944.

[3] E. SCHRÖDINGER, "Zur Theorie der Fall- und Steigversuche an Teilchen mit Brownscher Bewegung", *Physikalische Zeitschrift*, Vol. 16 (1915), pp. 289–295.

[4] A. WALD, "On cumulative sums of random variables", *Ann. Math. Stat.*, Vol. 15 (1944), pp. 283–296.

[5] M. C. K. TWEEDIE, "A mathematical investigation of some electrophoretic measurements on colloids," unpublished thesis for M.Sc. degree, University of Reading, England, 1941.

[6] M. C. K. TWEEDIE, "Functions of a statistical variate with given means, with special reference to Laplacian distributions," *Proc. Cambridge Philos. Soc.*, Vol. 43 (1947), pp. 41–49.

[7] E. J. SCOTT, *Transform Calculus with an Introduction to Complex Variables*, Harper, New York, 1955.

[8] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, 1941.

[9] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1950.