DEPARTMENT OF COMMERCE
SINCLAIR WEEKS, Secretary

WEATHER BUREAU
F. W. REICHELDERFER, Chief

MONTHLY WEATHER REVIEW

JAMES E. CASKEY, JR., Editor

Volume 86
Number 4

APRIL 1958

Closed June 15, 1958
Issued July 15, 1958

A NOTE ON THE GAMMA DISTRIBUTION

H. C. S. THOM

Office of Climatology, U. S. Weather Bureau, Washington, D. C.

[Manuscript received March 7, 1958; revised April 15, 1958]

ABSTRACT

The general properties of the gamma distribution, which has several applications in meteorology, are discussed. A short review of the general properties of good statistical estimators is given. This is applied to the gamma distribution to show that the maximum likelihood estimators are jointly sufficient. A new, simple approximation of the likelihood solutions is given, and the efficiency of the fitting procedure is computed.

1. INTRODUCTION

In 1947, the writer [1] developed approximate solutions of the maximum likelihood (M.L.) equations for the incomplete gamma distribution commonly called the gamma distribution. The purpose of this note is to give the development of the estimates which have since been widely employed. The first application of the methods was to rainfall data [2]; later it was found that the gamma distribution has wide application in meteorology to problems where the climatological variable has a physical lower bound of zero but no nonstatistical upper bound.

The gamma distribution is a special case of the Pearson Type III distribution where the locus parameter is zero. Fisher [3] first gave the M. L. equations for this distribution; however, as is often the case with M. L. estimation the equations are not conveniently solved. Our approximations make the estimation of the distribution parameters hardly more difficult than the estimation of the mean and standard deviation of the normal distribution.

2. THE GAMMA DISTRIBUTION AND PROPERTIES

The gamma distribution is a 2-parameter frequency distribution given by the equation

$$f(x)=\frac{1}{\beta^\gamma\Gamma(\gamma)}x^{\gamma-1}e^{-x/\beta};\quad \begin{matrix}\beta>0\\ \gamma>0\end{matrix}. \tag{1}$$

Here x is the random variable, β scales x and is therefore the scale parameter, γ is the shape parameter, Γ is the usual gamma function, and $f(x)=0$ for $x<0$. It will be noted that the distribution has a zero lower bound and is unlimited on the right. It is positively skewed, the amount of skew depending inversely on the shape factor γ. The mode of the distribution is at $\beta\ (\gamma-1)$ if $\gamma>1$ and at zero if $0<\gamma\leq1$. In the latter case, the distribution is J-shaped. For $\gamma=1$ the distribution is exponential with ordinate $1/\beta$ at $x=0$; for $\gamma<1$ the ordinate at $x=0$ is infinite. The gamma distribution is closely related to the chi square distribution, for $\chi^2/2$ is a gamma variate with $\gamma=\frac{1}{2}n$ and $\beta=1$.

The moments about zero of the gamma distribution are given by the relation

$$\mu_r'=\beta^r\gamma(\gamma+1)\ .\ .\ .\ (\gamma+r-1), \tag{2}$$

from which it follows immediately that the mean is

$$\mu_1'=\beta\gamma. \tag{3}$$

468423—58——1

From the moment relationships the second, third, and fourth moments about the mean are easily found to be

$$\mu_2=\sigma^2=\beta^2\gamma \quad (4)$$

$$\mu_3=2\,\beta^3\gamma \quad (5)$$

$$\mu_4=3\beta^4\gamma(\gamma+2). \quad (6)$$

Since the skewness statistic is $\sqrt{b_1}=\mu_3/\sigma^3$, we have from (4) and (5) that

$$\sqrt{b_1}=\frac{2}{\sqrt{\gamma}}. \quad (7)$$

Hence, the skewness goes to zero with increasing γ showing that the gamma distribution becomes symmetrical for large γ; in fact, it may be shown that the distribution approaches normality slowly as γ increases. For $\gamma>100$ it is approximately normal for climatological applications.

The main interest in applications to climatological analysis is not in equation (1) but in its integral which gives probability. This cannot be found except as an expansion in series or continued fractions. The integral from 0 to any value of the variate has been tabulated by Pearson [4], and this of course gives the probability that any value of the variate is less than the tabulated value.

Pearson used the moments for fitting the Type III frequency curve, so the arguments of his table are u and p. The variate u is scaled in terms of the standard deviation instead of β. Hence, $u=x/(\beta\sqrt{\gamma})$ or $x/\beta=u\sqrt{\gamma}$ in our notation. Also, his $p=\gamma-1$. To use Pearson's table we find $\gamma=p+1$ and multiply the u value by $\sqrt{\gamma}$. For certain purposes it may be more convenient to convert the x's to u values.

3. STATISTICAL ESTIMATORS

The main statistical problem in the application of the gamma distribution to climatological data is the estimation of the parameters β and γ from a sample record. The estimation problem is one of three basic problems in statistical analysis, and it will serve our present purpose to discuss the general problem briefly.

It has long been known that there are many ways to estimate the parameters in a statistical equation from a sample of data. Two of the more common methods are least squares and moments. It was found by Fisher [3] that the various methods of estimation do not give equally good results in the sense that some estimates or statistics are more variable than others. Clearly, the best estimates are those which have the smallest variability. For example, in samples of 10 from a normal population the mean or expected value could be estimated by averaging the smallest and the largest value, or it could be estimated by averaging all the observations. Obviously, the latter statistic using all the observations should be better than that using only two of the observations. In fact, it has been shown that the variability from sample to sample for sample size 10 as measured by the variance is twice as large when only the extreme observations are averaged. The median, which is also an estimate of mean, has a variance about one-third greater than the mean for sample size 10. From this it may be inferred that if we use the mean range as an estimate, we in effect discard half of our data; if we use the median we discard one-third of it. In climatology, where data are scarce, the use of inefficient estimators is clearly to be avoided.

Fisher [3] made a remarkable contribution to statistical analysis by developing a method of estimation originally due to Gauss which he called the method of maximum likelihood (M. L.). This method consists of maximizing what he calls the likelihood or the product of the frequency functions of a sample. If $f(x;\beta,\gamma)$ is any frequency function, the likelihood is defined to be

$$M=\prod_{i=1}^{n} f(x_i;\beta,\gamma). \quad (8)$$

where x_i is the ith value in a sample of n. To maximize this it is simplest to take logarithms before differentiating and setting to zero. This gives

$$L=\sum_{i=1}^{n} \log f(x_i;\beta,\gamma). \quad (9)$$

Differentiating partially with respect to β and γ gives the M. L. differential equations

$$\left.\begin{aligned}\frac{\partial L}{\partial\beta}&=0,\\ \frac{\partial L}{\partial\gamma}&=0,\end{aligned}\right\} \quad (10)$$

Solving these gives the M. L. estimates commonly written as $\hat{\beta}$ and $\hat{\gamma}$. The M. L. estimates have certain remarkable advantages not always possessed by other estimates which will now be discussed.

In order to assess the quality of estimators in general, Fisher defined three desirable properties of statistics; viz., consistency, efficiency, and sufficiency. These may be defined as follows:

1. If an estimator or statistic is consistent it converges in probability to its population or parameter value. This may be expressed in symbols by

$$P(|T_n-\theta|<\epsilon)>1-\eta;\ n>N. \quad (11)$$

T_n is an estimate of the parameter θ based on sample size n, ϵ and η are arbitrarily small quantities, and N is any integer. This means that $T_n=\theta$ when T_n is calculated from the whole population.

2. A consistent estimate T_1 is said to be more efficient than another consistent estimate T_2 if $v(T_1)<v(T_2)$; i. e., if the variance of T_1 is less than the variance T_2. An estimate is said to be efficient if it has the smallest variance of a class of consistent estimates. The efficiency of an estimate is defined as $v(\hat{T})/v(T)$ where $\hat{T}$ is M. L. estimate.

3. An estimate T is said to be sufficient if it exhausts all possible information on θ from a sample of any size. If T_1 and T_2 are two different estimates of θ not functionally related, an estimate T_1 of θ is sufficient if the joint distribution of T_1 and T_2 has the form

$$f=f_1(T_1,\ \theta)f_2(T_2|T_1) \tag{12}$$

where f_1 is the frequency distribution of T_1 and f_2 is the distribution of T_2 given a sample value of T_1. Once T_1 is known the probability of any range of values for T_2 is the same for all θ; hence, T_2 cannot give any information on θ which is not already available from T_1. Sufficiency is the most desirable property of an estimate, and such estimates are said to be optimum.

The superiority of M. L. estimates was demonstrated by Fisher and others when they proved that M. L. estimates are consistent and efficient and if a sufficient estimate exists, it will be given by the M. L. method.

4. MAXIMUM LIKELIHOOD ESTIMATES

Applying (9) to the gamma distribution equation (1) gives

$$L=-n\gamma \log \beta-n\log\Gamma(\gamma)+(\gamma-1)\Sigma \log x-\frac{1}{\beta}\Sigma x \tag{13}$$

where the summation is over the n sample values. Differentiating as indicated in equations (10) we find the M. L. equations

$$\bar{x}/\hat{\beta}-\hat{\gamma}=0 \tag{14}$$

$$\log \hat{\beta}+\frac{\partial}{\partial\hat{\gamma}} \log \Gamma(\hat{\gamma})-\frac{1}{n}\Sigma \log x=0. \tag{15}$$

Since $\frac{\partial}{\partial\hat{\gamma}} \log \Gamma(\hat{\gamma})$ is the digamma function, $\psi(\hat{\gamma})$, we may write (15) in the simplified form

$$\log \hat{\beta}+\psi(\hat{\gamma})-\frac{1}{n}\Sigma \log x=0. \tag{16}$$

Taking logarithms of (14) and substituting for $\log \hat{\beta}$ in (16) gives

$$\log \hat{\gamma}-\psi(\hat{\gamma})= \log \bar{x}-\frac{1}{n}\Sigma \log x. \tag{17}$$

This equation is implicit in $\hat{\gamma}$ but may be solved with some difficulty using the Davis [5] tables of the ψ-functions. Masuyama and Kuroiwa [6] prepared tables of $\log \hat{\gamma}-\psi(\hat{\gamma})$ from tables of logarithms and tables of the digamma functions.

We developed the application of the gamma distribution to precipitation before Masuyama and Kuroiwa's tables were available although, of course, we had also followed the equivalent procedure of using the Davis tables. To simplify the technique of fitting we developed an approximation to $\log \hat{\gamma}-\psi(\hat{\gamma})$ as follows: Nörlund [7] shows that

$$\psi(\gamma)= \log \gamma-1/(2\gamma)-\sum_{k=1}^{m}(-1)^{k-1}B_k/(2k\gamma^{2k})+R_m \tag{18}$$

is an asymptotic expansion in which B_k are the Bernoulli numbers, $B_1=1/6$, $B_2=1/30$, etc., and R_m is the remainder after m terms. For $\gamma\geq 1$ we may write the inequality

$$|R_m|<\frac{B_{m+1}}{(2m+2)\gamma^{2m+2}}. \tag{19}$$

For only $m=1$ and $\gamma=1$, $|R_m|<0.00833$ which is less than 1.5 percent of the table value $\psi(1)=-0.57722$ given by Davis [5]. The approximation, of course, increases in accuracy with γ. At $\gamma=2$ it is within 0.1 percent of table value. We are not, however, interested in approximating ψ but in approximating γ.

From (18) for $m=1$ we find

$$\psi(\gamma)=\log \gamma-1/(2\gamma)-1/(12\gamma^2). \tag{20}$$

Substituting in (17) we find

$$12\left(\log \bar{x}-\frac{1}{n}\Sigma \log x\right)\hat{\gamma}^2-6\hat{\gamma}-1=0.$$

Simplifying by letting $A=\log \bar{x}-\frac{1}{n}\Sigma \log x$ we have

$$12\,A\hat{\gamma}^2-6\hat{\gamma}-1=0, \tag{21}$$

which is a quadratic equation whose only pertinent root is

$$\hat{\gamma}=\frac{1+\sqrt{1+4A/3}}{4A}. \tag{22}$$

This together with equation (14) gives the M. L. estimates for the gamma distribution. It is only necessary to sum the natural logarithms of x and take the natural logarithm of the mean of x to provide the basic data for equations (14) and (21). Common logarithms may, of course, also be used by multiplying by the proper conversion factor.

The error in $\hat{\gamma}$ resulting from using only one term of equation (18) is not readily expressed in mathematical form; hence, we have computed the following table for correcting the estimate obtained from equation (22).

$\hat{\gamma}$	$\Delta\hat{\gamma}$	$\hat{\gamma}$	$\Delta\hat{\gamma}$	$\hat{\gamma}$	$\Delta\hat{\gamma}$	$\hat{\gamma}$	$\Delta\hat{\gamma}$
0.2	0.034	0.8	0.012	1.4	0.006	2.2	0.003
0.3	.029	0.9	.011	1.5	.005	2.3	.002
0.4	.025	1.0	.009	1.6	.005	3.1	.002
0.5	.021	1.1	.008	1.7	.004	3.2	.001
0.6	.017	1.2	.007	1.8	.004	5.5	.001
0.7	.014	1.3	.006	1.9	.003	5.6	.000

The value of $\Delta\hat{\gamma}$ is to be subtracted from the value of $\hat{\gamma}$ obtained from equation (22).

5. SUFFICIENCY OF THE ESTIMATES

Koopman [8] has given the necessary and sufficient conditions for a set of estimators to be jointly sufficient in the form

$$\log f=\sum_{k=1}^{p} A_k X_k+B+Y. \tag{23}$$

Here A_k and B are functions of the parameters of the distribution (β and γ in this case) and X_k and Y are functions of x. Taking logarithms of (1) we have

$$\log f=-x/\beta+(\gamma-1)\log x-\log\Gamma(\gamma)-\gamma\log\beta. \tag{24}$$

This is of the form (23) where $A_1=-1/\beta$, $A_2=\gamma-1$, $X_1=x$, $X_2=\log x$, $Y=0$, and $B=-(\log\Gamma(\gamma)+\gamma\log\beta)$. This shows that $\hat{\beta}$ and $\hat{\gamma}$ are jointly sufficient estimates of β and γ. Thus, no other estimates of β and γ can give more information on these parameters, and $\hat{\beta}$ and $\hat{\gamma}$ are optimum statistics and have, indeed, a highly desirable property. Fisher [3] has also shown that $\bar{x}$ is a 100 percent efficient estimate of the population mean. This is important in many applications of the gamma distribution.

6. VARIANCE OF THE ESTIMATORS

Fisher [3] has shown how to obtain the large sample variance of M. L. estimates in general and has demonstrated that M. L. estimates are normally distributed in large samples. Using matrix methods we define

$$\sigma^{ij}=-E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f\right], \tag{25}$$

where σ^{ij} is the ijth element of a matrix, the θ's are the parameters, and E is the expected value operator. If further we define the matrix equation

$$[\sigma_{ij}]=[\sigma^{ij}]^{-1} \tag{26}$$

the variance-covariance matrix is

$$C=\left[\frac{1}{n}\sigma_{ij}\right]. \tag{27}$$

Since θ takes on two values β and γ, we may write the second partial derivatives as L_{11}, L_{22}, and L_{12} where the subscripts refer to the parameters. Differentiating equation (24) partially with respect to β and γ we find

$$L_{11}=\frac{\gamma}{\beta^2}-\frac{2x}{\beta^3}, \tag{28}$$

$$L_{22}=-\psi'(\gamma), \tag{29}$$

$$L_{12}=L_{21}=-\frac{1}{\beta}, \tag{30}$$

where ψ' is the trigamma function.

We have seen above in equation (3) that $E(x)=\beta\gamma$; hence

$$\sigma^{11}=\frac{\gamma}{\beta^2}. \tag{31}$$

Since (29) and (30) do not involve the random variable x, we have simply

$$\sigma^{22}=\psi'(\gamma) \tag{32}$$

$$\sigma^{12}=\frac{1}{\beta}=\sigma^{21}. \tag{33}$$

We now may write the matrix

$$[\sigma^{ij}]=\begin{bmatrix}\frac{\gamma}{\beta^2} & \frac{1}{\beta}\\ \frac{1}{\beta} & \psi'(\gamma)\end{bmatrix}. \tag{34}$$

The inverse of this after multiplying each term by $1/n$ as required by (27) is

$$\left[\frac{1}{n}\sigma_{ij}\right]=\begin{bmatrix}\frac{\beta^2\psi'(\gamma)}{n(\gamma\psi'(\gamma)-1)} & \frac{-\beta}{n(\gamma\psi'(\gamma)-1)}\\ \frac{-\beta}{n(\gamma\psi'(\gamma)-1)} & \frac{\gamma}{n(\gamma\psi'(\gamma)-1)}\end{bmatrix} \tag{35}$$

which is the variance-covariance matrix. From this it immediately follows that the variances and covariance are

$$v(\hat{\beta})=\frac{\beta^2\psi'(\gamma)}{n(\gamma\psi'(\gamma)-1)} \tag{36}$$

$$v(\hat{\gamma})=\frac{\gamma}{n(\gamma\psi'(\gamma)-1)} \tag{37}$$

$$\text{cov}\,(\hat{\beta},\hat{\gamma})=\frac{-\beta}{n(\gamma\psi'(\gamma)-1)} \tag{38}$$

Since $\text{cov}\,(\hat{\beta},\hat{\gamma})=r\sqrt{v(\hat{\beta})v(\hat{\gamma})}$, we easily find the correlation r between $\hat{\beta}$ and $\hat{\gamma}$ to be

$$r=\frac{-1}{\sqrt{\gamma\psi'(\gamma)}}. \tag{39}$$

This is negative as it is seen it must be by equation (14), for if the mean is fixed, $\hat{\beta}$ and $\hat{\gamma}$ must vary inversely. The large sample variances were also given by Masuyama and Kuroiwa.

7. EFFICIENCY OF THE METHOD OF MOMENTS

Although the method of moments is widely used in fitting frequency distributions in climatology, it is the exceptional case when this method proves to be fully efficient in estimating climatological parameters. The most prominent exception is, of course, the normal distribution where the moment estimates are jointly sufficient and so are identical with the M. L. estimates. As we have seen the M. L. method of estimation is always superior to moments when they give different results.

It is of interest then to compute the efficiency of the moment estimates for the gamma distribution to evaluate the superiority of M. L. estimates. We have seen that the efficiency of a moment estimate is the ratio of the variance of the M. L. estimate to that of the moment estimate. The M. L. variances are given by equations (36) and (37) and the variances of the moment estimates may be worked out as follows:

$$\beta_m = s^2/\bar{x} \tag{40}$$

$$\gamma_m = \bar{x}^2/s^2. \tag{41}$$

The problem now is to find the variance of these two estimates. This may be done using Taylor's formula. Taking the total differential of (40) with respect to $\bar{x}$ and s^2 gives

$$\Delta\beta_m = \frac{1}{\bar{x}}\,\Delta s^2 - \frac{s^2}{\bar{x}^2}\,\Delta\bar{x} \tag{42}$$

in which the coefficients of the delta increments are the partial derivatives and the increments of higher order are ignored. Squaring and taking expected values we find

$$E(\Delta\beta_m)^2 = \frac{1}{\bar{x}^2}\,E(\Delta s^2)^2 + \frac{s^4}{\bar{x}^4}\,E(\Delta\bar{x})^2 - \frac{2s^2}{\bar{x}^3}\,E(\Delta\bar{x}\Delta s^2). \tag{43}$$

Since the expected value of the square of an error is the variance, and the expected value of an error product is the covariance, we have

$$v(\beta_m) = \frac{1}{\bar{x}^2}\,v(s^2) + \frac{s^4}{\bar{x}^4}\,v(\bar{x}) - \frac{2s^2}{\bar{x}^3}\,\mathrm{cov}\,(\bar{x}, s^2). \tag{44}$$

Now the variances of s^2, $\bar{x}$, and cov $(\bar{x}, s^2)$ are known in large samples to be [9]

$$v(s^2) = (\mu_4 - \mu_2^2)/n \tag{45}$$

$$v(\bar{x}) = \mu_2/n \tag{46}$$

$$\mathrm{cov}\,(\bar{x}, s^2) = (n-1)\,\mu_3/n^2. \tag{47}$$

Substituting values of the moments of the gamma distribution given by equations (3), (4), (5), and (6) for the moments in (45), (46), and (47) and then for the sample moments, variances, and covariance in (44) we find after some simplification

$$v(\beta_m) = \frac{\beta^2}{n\gamma}\,[2\gamma + 7 - 4(n-1)/n] \tag{48}$$

which for large n becomes

$$v(\beta_m) = \frac{\beta^2}{n\gamma}\,(2\gamma + 3). \tag{49}$$

Performing similar operations on equation (41) we find for large samples

$$v(\gamma_m) = \frac{2\gamma}{n}\,(\gamma + 1). \tag{50}$$

If we assume $\beta = 1$, a common value for climatological data, we find for $\gamma = 1$

$$v(\beta_m) = 5/n;\;\; v(\gamma_m) = 4/n. \tag{51}$$

and for $\gamma = 10$,

$$v(\beta_m) = 2.3/n;\;\; v\,(\gamma_m) = 220/n. \tag{52}$$

Using the Davis tables [5] we find the M. L. variances from (36) and (37) for $\gamma = 1$

$$v(\hat{\beta}) = 2.55/n;\;\; v\,(\hat{\gamma}) = 1.55/n \tag{53}$$

and for $\gamma = 10$,

$$v(\hat{\beta}) = 2.04/n;\;\; v\,(\hat{\gamma}) = 193.6/n. \tag{54}$$

According to our definition of efficiency we must take the ratio of the variance of the M. L. estimate to the moment estimate. This gives from equation (51) and (53) for $\gamma = 1$,

Eff $(\beta_m) = 51$ percent

Eff $(\gamma_m) = 39$ percent

and for $\gamma = 10$

Eff $(\beta_m) = 89$ percent

Eff $(\gamma_m) = 88$ percent

We see that for $\gamma < 10$ the method of moments produces unacceptable estimates for both β and γ. For γ near 1 the moment estimates use only 50 percent of the information in the sample for estimating β and only 40 percent for γ. Thus, for β the M. L. estimator would do as well with half the length of record as the moment estimate and for γ with two-fifths the record length. Hence, use of the moment estimates in effect results in discarding half the record in estimating β and three-fifths of the record in estimating γ. For $\gamma = 10$ the efficiencies of β_m and γ_m both approach satisfactory values. In view of the difficulty in obtaining homogeneous climatological data series, it seems at least a questionable procedure to employ inefficient estimators which do not make the best use of the available climatological samples.

REFERENCES

1. H. C. S. Thom, "A Note on the Gamma Distribution" Statistical Laboratory, Iowa State College, 1947 (Manuscript).
2. G. L. Barger and H. C. S. Thom, "Evaluation of Drought Hazard," *Agronomy Journal* vol. 41, No. 11, Nov. 1949, pp. 519–526
3. R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions Royal Society*, Series A, vol. 222, 1941, pp. 309–368.
4. K. Pearson et al., *Tables of the Incomplete Γ-function*, Cambridge University Press. Reissue 1951.

5. H. T. Davis, *Tables of Higher Mathematical Functions*, vol. I and II, Principia Press, Bloomington, 1933.
6. M. Masuyama and Y. Kuroiwa, "Table of the Likelihood Solutions of Gamma Distribution and its Medical Applications," *Statistical Application Research*, vol. 1, No. 1, 1951.
7. N. E. Nörlund, *Vorlesungen über Differenzenrechnung*, Springer, Berlin, 1924, p. 101, 106.
8. B. O. Koopman, "Distributions Admitting a Sufficient Statistic", *Transactions, American Mathematical Society*, vol. 39, 1936, p. 399.
9. H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 348.