

A PHYSICAL BASIS FOR THE GENERALIZED GAMMA DISTRIBUTION*

By JOHN H. LIENHARD AND PAUL L. MEYER (*Washington State University*)

Introduction. A number of known families of probability distributions can be derived from requirements that are physical in the sense that they describe the random behavior of the event under consideration. The Poisson process is typical of these: If $X(t)$ is equal to the number of occurrences of a specified event in the interval $[0, t)$, then one can show that $X(t)$ has, for each t , a Poisson distribution, if the probabilistic behavior of the event satisfies a few very simple physical requirements [1]. The normal distribution may also be derived from a few physical requirements.

In this note we shall employ a simple model and statistical-mechanical methods to derive the three-parameter generalized gamma distribution. The history of this family of distributions was reviewed and further properties were discussed in 1962 by Stacy [2]. Subsequent work on statistical problems associated with the distribution has been done by Bain and Weeks [3]. Special cases of the generalized gamma distribution include the Weibull, gamma, Rayleigh, exponential and Maxwell velocity distributions. Another special case is a distribution recently derived on a statistical-mechanical basis to describe rainfall run-off from a watershed [4].

The model. We shall consider the following situation: The occurrence of an event, such as the failure of a component or system, depends on some variable such as the stress to which the part has been subjected or the time during which it has been subjected to a given level of stress or use. This variable (be it stress or time) will be designated by t and the number of occurrences of the event during the interval $[t_{i-1}, t_i)$ will be designated by N_i , where $t_i - t_{i-1} \equiv \Delta t$ and t_0 is the arbitrary origin.** The requirements that we shall impose upon the N_i 's are as follows:

1. The total number of occurrences of the event is fixed:

$$\sum_{i=1}^{\infty} N_i = N. \quad (1)$$

The N_i 's and N are assumed to be large numbers.

2. For each choice of β , the following sum is a positive constant:

$$\sum_{i=1}^{\infty} (N_i/N) t_i^\beta = K. \quad (2)$$

3. The number of distinguishable ways, g_i , in which the event can occur in the interval $[t_{i-1}, t_i)$ is proportional to a specified power of t_i . Thus:

$$g_i = A t_i^{\alpha-1}. \quad (3)$$

We shall restrict our subsequent considerations to α , β , and $K > 0$.

The derivation of the distribution of the number of events will now consist of determining the most probable distribution satisfying these requirements. The derivation, as it happens, does not differ in principle from the derivation of the velocity distribution in a degenerate gas, using Boltzmann statistics (see, e.g., [5]).

*Received March 1, 1966; revised manuscript received May 19, 1966.

**Without loss of generality, t_0 can be taken as equal to zero.

Derivation of the probability density function. We designate as W the number of ways in which N distinguishable occurrences of the event can take place, if N_i of these occurrences must take place in $[t_{i-1}, t_i)$ and if the number of ways the event can occur in $[t_{i-1}, t_i)$ is given by g_i . It is well known [5] that:

$$W = N! \prod_{i=1}^{\infty} (g_i^{N_i})/N_i! \quad (4)$$

Let us denote by \bar{N}_i , those values of N_i which maximize W , subject to the requirements (1), (2), and (3) of the previous section. The following two results may now be established:

(i) The explicit expression for \bar{N}_i is given by

$$\frac{\bar{N}_i}{N} = \frac{\Delta t [\beta(\beta K/\alpha)^{-\alpha/\beta}]}{\Gamma(\alpha/\beta)} t_i^{\alpha-1} \exp \left[-\frac{\alpha}{\beta} \frac{t_i^\beta}{K} \right] \quad (5)$$

which, we note, involves α , β , and K , but not the constant of proportionality, A .

(ii) The maximum value of W (corresponding to $N_i = \bar{N}_i$) is very much larger than the value of W corresponding to N_i 's that are significantly different from the \bar{N}_i 's, provided the N_i 's and N are large.

The proof of assertion (i) is given in the Appendix. The argument leading to (ii) is well known from statistical mechanics and may be found, for example, in [5].

We may thus suppose that the \bar{N}_i/N 's represent the discrete probability distribution associated with the random variable T , where T is the time or stress at which the first occurrence of the event under consideration takes place. Accordingly:

$$P(t_{i-1} \leq T < t_i] = \bar{N}_i/N, \quad i = 1, 2, \dots \quad (6)$$

We shall now approximate this discrete distribution with a continuous probability density function, f , as follows. From Eq. (6) we can write:

$$\bar{N}_i/N = \int_{t_{i-1}}^{t_i} f(t) dt. \quad (7)$$

In accordance with the mean value theorem, the integral in Eq. (7) becomes $\Delta t f(\xi)$, where $(t_i - \Delta t) \leq \xi \leq t_i$. Hence, letting $\Delta t \rightarrow 0$ and using Eq. (5) for \bar{N}_i/N , we obtain:

$$f(t) = \left[\frac{\beta}{\Gamma(\alpha/\beta)} \left(\frac{\alpha}{\beta K} \right)^{\alpha/\beta} \right] t^{\alpha-1} \exp \left[-\frac{\alpha}{\beta} \frac{t^\beta}{K} \right], \quad t \geq 0, \quad (8)$$

where α , β , and K should be > 0 so that $f(t) \geq 0$. An easy integration shows that $\int_0^\infty f(t) dt = 1$, for all permissible choices of α , β , and K .

The introduction of $a \equiv (\beta K/\alpha)^{1/\beta}$ into Eq. (8) yields:

$$f(t) = \left[\frac{\beta}{a^\alpha \Gamma(\alpha/\beta)} \right] t^{\alpha-1} \exp [-(t/a)^\beta] \quad (8a)$$

which is the form in which the distribution was given by Stacy [2].

Properties of the family of distributions. It is easily verified that the expectation of T^β , $E(T^\beta) = \int_0^\infty t^\beta f(t) dt = K$. This is, of course, as it should be in view of the constraint (2) and it indicates that in the continuous model, as represented by the probability density function, f , this constraint still holds.

By making the following change of scale, the three-parameter family of distributions as given by Eq. (8) may be reduced to a two-parameter family:

Let $S = T/K^{1/\beta}$ and denote by $h(s)$ the probability density function of S . It is easy to show that:

$$h(s) = \left[\frac{\beta}{\Gamma(\alpha/\beta)} \left(\frac{\alpha}{\beta} \right)^{\alpha/\beta} \right] s^{\alpha-1} \exp \left[-\frac{\alpha}{\beta} s^\beta \right], \quad s \geq 0.$$

Although S depends on K , its probability distribution does not. The distribution of S is normalized in the sense that $E(S^\beta) = 1$.

A number of familiar distributions, many of which were noted previously [2, 3], can be obtained as special cases of Eq. (8) by making certain choices for the parameters α and β . The parameter K will appear in all of the distributions discussed below but it could be removed by the normalization suggested above.

a. *The Weibull distribution:* $\beta = \alpha$.

$$f(t) = \frac{\alpha}{K} t^{\alpha-1} \exp \left[-\frac{t^\alpha}{K} \right].$$

It is helpful to relate this result to the requirements upon which it is based. Failures of a system that can fail in $At^{\alpha-1}$ ways at time or stress, t , will be distributed in accordance with Weibull's Law.*

b. *The hydrograph distribution:* $\beta = 2$.

$$f(t) = \left[\frac{2}{\Gamma(\alpha/2)} \left(\frac{\alpha}{2K} \right)^{\alpha/2} \right] t^{\alpha-1} \exp \left[-\frac{\alpha}{2} \frac{t^2}{K} \right].$$

This distribution was originally formulated to describe the distribution of rainfall run-off from a watershed, as a function of time after a sudden storm. In this case, g_i was the number of ways in which a raindrop could find its way to the gaging station after time t_i . Long slender watersheds and fan-shaped watersheds were characterized by $\alpha = 2$ and $\alpha = 3$, respectively.

c. *The gamma distribution:* $\beta = 1$.

$$f(t) = \left[\frac{(\alpha/K)^\alpha}{\Gamma(\alpha)} \right] t^{\alpha-1} \exp [-\alpha(t/K)].$$

d. *The Rayleigh distribution:* $\beta = \alpha = 2$.

$$f(t) = \frac{2}{K} t \exp [-t^2/K].$$

This is a special case of the Weibull distribution (a), and one of the special cases of the hydrograph distribution (b), mentioned above.

e. *The Maxwell molecular speed distribution:* $\beta = 2, \alpha = 3$.

$$f(t) = \left[\frac{(54/\pi)^{1/2}}{K^{3/2}} \right] t^2 \exp \left[-\frac{3}{2} \frac{t^2}{K} \right].$$

*It is very interesting to note that the Weibull distribution can also be derived on the basis of a failure rate which is proportional to $t^{\alpha-1}$. Of course the physical meaning of g_i differs considerably from that of the failure rate.

This well known law is also one of the special cases of distribution (b) mentioned above.

f. *The Maxwell molecular velocity distribution:* $\beta = 2, \alpha = 1$.

$$f(t) = (2/K\pi)^{1/2} \exp [-t^2/2K].$$

This distribution was the first one to be obtained by the statistical-mechanical procedure outlined here. The distinction between distributions (e) and (f) is explored in [6].

g. *The exponential distribution:* $\beta = \alpha = 1$.

$$f(t) = \frac{1}{K} \exp [-t/K].$$

This is another special case of the Weibull distribution.

Conclusion. The generalized gamma distribution has been obtained by applying a statistical-mechanical method to a physical model. The parameters, K , α , and β , are associated with the scale of the distribution, the number of ways in which the event can occur, and a moment of the distribution, respectively. By varying the parameters one can obtain a large number of probabilistic models for the description of random phenomena. The following considerations should be weighed in selecting the three parameters for any application.

(1) The parameter K always establishes the scale of the distribution.

(2) The parameter α reflects the physical process which leads to the event. It is our hope that (on the basis of the present model) experimentally determined values of α might be used to infer details about the occurrence of the event. Such application might be especially fruitful in connection with the Weibull distribution.

(3) The choice of β is of secondary importance in terms of the physical model. Thus, for example, the Weibull distribution (for which β is chosen equal to α) and the hydrograph distribution (for which β is chosen equal to 2) might well be used interchangeably to represent data.

APPENDIX

We maximize W by determining the set of N_i 's for which $\ln W$ is maximum. Thus, after substituting Eq. (3) into Eq. (4), and using Stirling's approximation we obtain:

$$d \ln W = \sum_{i=1}^{\infty} [\ln A t_i^{\alpha-1} - \ln N_i] dN_i = 0.$$

The constraints (1) and (2) give, in differential form,

$$-\sum_{i=1}^{\infty} \rho dN_i = 0 \quad \text{and} \quad -\sum_{i=1}^{\infty} \gamma t_i^{\beta} dN_i = 0,$$

where ρ and γ are Lagrangian multipliers. Adding these three equations we get:

$$\sum_{i=1}^{\infty} [\ln (A t_i^{\alpha-1} / N_i) - \rho - \gamma t_i^{\beta}] dN_i = 0.$$

The coefficients of dN_i must vanish identically, so:

$$N_i = (A e^{-\rho}) [t_i^{\alpha-1} \exp (-\gamma t_i^{\beta})]$$

or:

$$\frac{N_i}{N} = \left[\sum_{i=1}^{\infty} t_i^{\alpha-1} \exp(-\gamma t_i^\beta) \right]^{-1} t_i^{\alpha-1} \exp(-\gamma t_i^\beta). \quad (9)$$

The summations in the bracketed coefficient (above) and in the two constraints can be replaced with definite integrals, as Δt is taken to be small. Accordingly the coefficient becomes: $[\beta(\gamma \Delta t^\beta)^{\alpha/\beta} / \Delta t^{\alpha-1} \Gamma(\alpha/\beta)]$ and elimination of $N/Ae^{-\rho}$ from the two constraints gives: $\gamma = \alpha/\beta K$. If we denote N_i 's for which W is maximum as \bar{N}_i , Eq. (9) becomes:

$$\frac{\bar{N}_i}{N} = \Delta t \left[\frac{\beta[\beta K/\alpha]^{-\alpha/\beta}}{\Gamma(\alpha/\beta)} \right] t_i^{\alpha-1} \exp \left[-\frac{\alpha}{\beta} \frac{t_i^\beta}{K} \right]$$

which is eq. (5).

REFERENCES

- [1] P. L. Meyer, *Introductory probability and statistical applications*, Addison-Wesley, Reading, Mass., 1965
- [2] E. W. Stacy, *A generalization of the gamma distribution*, Ann. Math. Statist., **33**, 1187 (1962)
- [3] L. J. Bain and D. L. Weeks, *Tolerance limits for the generalized gamma distribution*. J. Amer. Statist. Assoc., **60**, 1142 (1965)
- [4] J. H. Lienhard, *A statistical mechanical prediction of the dimensionless unit hydrograph*, J. Geophys. Res., **69**, no. 24, 5231 (1964)
- [5] A. Sommerfeld, *Thermodynamics and statistical mechanics*, Lectures on Theoretical Physics, Vol. 5, Academic Press, New York, 1956, pp. 213-220
- [6] S. Chapman and T. G. Cowling, *The mathematical theory of nonuniform gases*, Cambridge Univ. Press, London, 1960, Secs. 4.10 and 4.11