GENERALIZED POISSON DISTRIBUTION

By F. E. Satterthwaite
Aetna Life Insurance Company

1. Introduction. The Poisson distribution is one of the most fundamental of statistical distributions. It is the distribution law for the number of events if the probability of an event happening in any infinitesimal unit of time is independent of the probability of its happening in any other unit of time. Frequently when we analyze statistics which obey the Poisson law it is desirable to give varying weights to the different events instead of considering them all of equal value. Such is the case in analyzing insurance statistics where the events are the claims received by the office and the weights are the cost of the claim to the company. We shall now show how the Poisson distribution can be generalized so as to be adequate for such an analysis.

2. First development. Let $f(x, \alpha)$ be the distribution function of the weights assigned to the events where the variable, x, refers to the weight and the variable, α, refers to time. The characteristic function of $f(x, \alpha)$ is

$$\phi(t, \alpha) = \int e^{itx} f(x, \alpha)\, dx.$$

Also let $p(\alpha)\, d\alpha$ be the probability that an event will occur in the infinitesimal unit of time, α to $\alpha + d\alpha$. If y represents the sum of the weights, the distribution function of y for this unit of time is

$$\begin{aligned} F_{d\alpha}(y, \alpha) &= 1 - p(\alpha)\, d\alpha, & y &= 0 \\ &= f(y, \alpha) p(\alpha)\, d\alpha, & y &> 0. \end{aligned} \tag{1}$$

The characteristic function of this distribution is

$$\begin{aligned} \Phi_{d\alpha}(t, \alpha) &= e^{it0}(1 - p(\alpha)\, d\alpha) + p(\alpha)\, d\alpha \int e^{ity} f(y, \alpha)\, dy \\ &= 1 - p(\alpha)\, d\alpha(1 - \phi(t, \alpha)) \\ &= e^{-p(\alpha) d\alpha (1 - \phi(t, \alpha))}. \end{aligned} \tag{2}$$

In forming equations (1) and (2) we ignore infinitesimals of orders higher than the first in the $d\alpha$.

The expected number of events in the period of time from α_1 to α_2 is

$$P = \int_{\alpha_1}^{\alpha_2} p(\alpha)\, d\alpha,$$

and the mean distribution of weights during the same period of time is

$$f(x) = \int_{\alpha_1}^{\alpha_2} [p(\alpha)/P] f(x, \alpha)\, d\alpha.$$

The characteristic function of this mean distribution of weights is

$$\phi(t) = \int e^{itx} f(x)\, dx$$
$$= \int [p(\alpha)/P]\phi(t, \alpha)\, d\alpha.$$

These equations are based on the assumption that the probability of an event occurring in any unit of time is independent of the probability of its occurrence in any other unit of time and also the assumption that the weights assigned to each event are independent. These assumptions are implied in all that follows.

Since the characteristic function of the sum of independent variables is equal to the product of the respective characteristic functions, the characteristic function of the sum of the weights during the period of time, α_1 to α_2, is

$$\Phi(t) = \Pi \Phi_{d\alpha}(t, \alpha)$$
$$(3) \qquad = e^{-\int p(\alpha)d\alpha + \int p(\alpha)\phi(t,\alpha)d\alpha}$$
$$= e^{-P(1-\phi(t))}.$$

Applying the Fourier transformation, the distribution function of the sum of the weights is

$$F(y) = \frac{1}{2\pi} \int e^{-ity-P(1-\phi(t))}\, dt.$$

Equation (3) gives a convenient method for defining a generalized Poisson distribution. Any distribution which has a characteristic function in the form of $\Phi(t)$ where $\phi(t)$ is the characteristic function of an arbitrary distribution will have all the properties of a generalized Poisson distribution.

3. Second development. If we let $\phi(t)$ represent the characteristic function of an arbitrary distribution, the characteristic function of the sum of n independent items obeying such a distribution law is $\Phi_n(t) = [\phi(t)]^n$. If instead of considering n to be a fixed quantity we assume that it is an independent statistical variable obeying the Poisson distribution law with mean P, the characteristic function of the sum, y, of the items of the sample becomes

$$\Phi(t) = \Sigma_n \frac{1}{n!} P^n[\phi(t)]^n e^{-P}$$
$$= e^{-P(1-\phi(t))}.$$

Therefore y is seen to obey the generalized Poisson distribution law.

4. Properties. The generalized Poisson distribution preserves the unique and very important property of the Poisson distribution that nowhere in its development is it necessary to make any assumptions regarding homogeneity. The only requirement is that the occurrence of and weight assigned to any event shall be independent of the occurrence of or weight assigned to any other event.

The distribution of the sum of the weights is a function of the expected number of events, P, and of the mean distribution of weights, $f(x)$, alone. It is independent of the way in which P and $f(x)$ are made up. Thus, if we are studying the distribution of the sum of the weights over a period of a year and if P and $f(x)$ vary with the seasons, the distribution of y is no different than it would be if P and $f(x)$ were constant. It is only necessary that the $f(x)$'s for the different seasons be weighted in proportion to the expected number of events in determining the mean $f(x)$.

Note also that in the first development it is not necessary that the variable, α, refer to time. It could just as well refer to different classes of events distinguished on any other basis. Therefore, heterogeneous material may be combined in an analysis if it is possible to determine the appropriate mean distribution of weights.

For a given weight distribution the generalized Poisson distribution for an expected number of events, nP, is identical with the distribution of the sum of n independent items each of which obeys a generalized Poisson distribution with P expected events.

Because of the property described in the preceding paragraph it is immediately apparent that a generalized Poisson distribution obeys the law of large numbers. As the number of expected events increases the distribution approaches the normal distribution.

5. Moments. The moments of a generalized Poisson distribution are functions of the moments of the underlying weight distribution. By differentiating the characteristic function we obtain the following formulas in which the presubscript, $_0$, refers to the moments of the weight distribution, $f(x)$:

$$\mu_1' = P\,{}_0\mu_1' = m$$

$$\mu_2 = P\,{}_0\mu_2' = \sigma^2$$

$$\mu_3 = P\,{}_0\mu_3'$$

$$\mu_4 = P\,{}_0\mu_4' + 3(P\,{}_0\mu_2')^2.$$

The above formulas may be verified through general reasoning by considering the moments of the distribution, $F_{d\alpha}(y, \alpha)$ (see equation (1)). This distribution refers to an infinitesimal unit of time and all the moments about zero are infinitesimals of the first order. In passing from the moments about zero to the moments about the mean the corrections are all infinitesimals of at least the second order. Therefore, the corrections may be ignored and the moments about the mean may be considered to be equal to those about zero. The above formulas follow if we take a sample of size $P/pd\alpha$ from this population.

In order to obtain Pearson's moment functions for a generalized Poisson distribution for any given mean value it is convenient to calculate the following parameters of the weight distribution:

$$(4)\qquad \begin{aligned} {}_0m &= {}_0\mu_1' \\ {}_0\sigma^2 &= {}_0\mu_2'/{}_0m \\ {}_0\beta_1 &= ({}_0\mu_3'/{}_0m)^2/{}_0\sigma^6 \\ {}_0(\beta_2 - 3) &= ({}_0\mu_4'/{}_0m)/{}_0\sigma^4. \end{aligned}$$

The Pearson moment functions then take the convenient forms:

$$(5)\qquad \begin{aligned} \sigma^2/m^2 &= {}_0\sigma^2/m \\ \beta_1 &= {}_0\beta_1/m \\ (\beta_2 - 3) &= {}_0(\beta_2 - 3)/m. \end{aligned}$$

6. Further generalizations. Often the expected number of events is not known but can be estimated to a greater or less degree of accuracy. In such a case it is convenient to assume that P is a statistical variable distributed about some expected value, say P'. A Type III distribution,

$$g(P) = \frac{1}{\Gamma(b)}\left(\frac{b}{P'}\right)^b P^{b-1} e^{-bP/P'},$$

will generally be as satisfactory as any to assume for P. The parameter, b, can be chosen to give any desired standard deviation. The characteristic function of the distribution of the sum of the weights under these conditions becomes

$$\begin{aligned} \Phi'(t) &= \int e^{-P(1-\phi(t))} g(P)\, dP \\ &= \left[1 + \frac{P'(1 - \phi(t))}{b}\right]^{-b}. \end{aligned}$$

The second development suggests another generalization. Instead of assuming that the number of events, n, is distributed in accord with the Poisson distribution, we may assume any discrete, non-negative distribution, $h(n)$. The distribution function for the sum of the weights is then

$$F'(y) = \Sigma h(n) f(y, n)$$

where $f(y, n)$ is the distribution function for the sum of n independent weights. The variance, σ^2, of this distribution is given by the formula,

$$\frac{\sigma^2}{m^2} = \frac{{}_n\sigma^2}{{}_nm^2} + \frac{1}{{}_nm}\frac{{}_0\sigma^2}{{}_0m^2},$$

where m refers to the mean, n refers to the distribution $h(n)$, and ${}_0$ refers to the weight distribution. Some writers have assumed that statistics of this type are distributed as a product. Such an assumption is incorrect and causes an overstatement of the variance to the amount of ${}_nm \cdot {}_0m^2 \cdot {}_n\sigma^2 \cdot {}_0\sigma^2$.

7. Application. In Table I is shown the distribution of claims under a certain plan of group sickness and accident insurance. The parameters, (4), for this distribution are

(6) $\quad {}_0m = 3.62, \quad {}_0\sigma^2 = 8.1, \quad {}_0\beta_1 = 14, \quad {}_0(\beta_2 - 3) = 15.$

This distribution is in terms of weeks per claim. The insurance company is interested in the financial cost per claim. A study shows that the distribution of the rate of weekly indemnity to which different classes of employees are entitled has the average parameters,

(7) $\quad {}_1m = 15.25, \quad {}_1\sigma^2 = 16.5, \quad {}_1\beta_1 = 20, \quad {}_1(\beta_2 - 3) = 25.$

Since the moment about zero of the product of independent statistics is equal to the product of the moments, it is permissible to multiply together the corre-

TABLE I

Nearest Duration of Claim in Weeks	Number of Claims per Year per 10,000 Employees
0	197
1	418
2	173
3	109
4	84
5	58
6	45
7	35
8	27
9	24
10	20
11	17
12	14
13	128

sponding parameters of (6) and (7) to obtain the average parameters for the distribution of the financial cost per claim. These are

$${}_2m = 55.2, \quad {}_2\sigma^2 = 134, \quad {}_2\beta_1 = 280, \quad {}_2(\beta_2 - 3) = 375.$$

In order to study the distribution of cost under a group of policies for each of which $180 in claims is expected, we apply equations (5) to obtain the parameters,

(8) $\quad \sigma^2/m^2 = .74, \quad \beta_1 = 1.6, \quad \beta_2 - 3 = 2.1.$

Since the expected number of claims is

$$P = 180/55.2 = 3.3$$

the probability that there will not be any claims under a policy is

$$h(0) = \frac{1}{0!}(3.3)^0 e^{-3.3} = .037.$$

Adjusting the parameters, (8), to remove the zero claims and choosing the scale so as to express the results as loss ratios gives the parameters,

$$m = 61.6\%, \qquad \sigma = 52.8\%, \qquad \beta_1 = 1.57, \qquad \beta_2 = 4.90.$$

A Pearson Type I curve fitted to these parameters intersects the axis well below the zero point. Therefore β_2 was reduced to 4.59 which gives the expected distribution shown in Table II.

Table II also shows the actual distribution of loss ratios experienced by one of the larger group insurance carriers under policies in this class. The Chi-square test for goodness of fit gives,

TABLE II

Experience under Group Sickness and Accident Insurance Policies

Ratio of Losses to Premiums	Number of Policies	
	Expected	Actual
0	18	11
.01– .09	47	37
.10– .19	53	45
.20– .29	50	56
.30– .39	45	38
.40– .49	41	47
.50– .59	36	39
.60– .69	32	41
.70– .79	28	37
.80– .89	24	20
.90– .99	21	29
1.00–1.19	32	30
1.20–1.39	23	22
1.40–1.59	17	22
1.60–1.99	19	14
2.00 and over	11	9

$$\chi^2 = 23, \qquad 14 \text{ degrees of freedom},$$

which corresponds to a probability of 5 per cent. Thus it is apparent that theory and experience are in fair agreement considering that no allowance was made for the lack of homogeneity "between policies." (This should not be confused with the homogeneity "within policies" covered in the theory.)

If the expected number of events is small, especially if the weight distribution is irregular or discrete, it is sometimes advisable to use the following method:

1. Use summation or approximate integration to obtain the distribution, $f(y, n)$, of the sum of n independent weights for $n = 1, 2, 3,$ and 4. The formula is

$$f(y, n+1) = \int_0^y f(x)f(y-x, n)\, dx.$$

2. Determine the generalized Poisson distribution for P, the expected number of events, equal to some small number, say $\frac{1}{4}$. The formula is

$$F(y, P) = \Sigma \frac{1}{n!} P^n e^{-P} f(y, n).$$

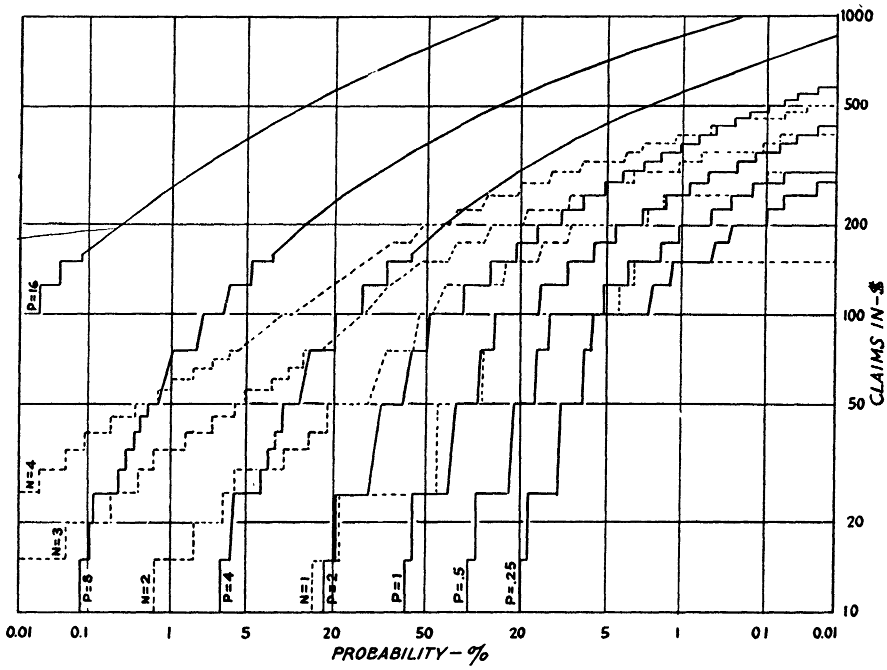

FIG. 1. Surgical Fee Insurance. ----, Distribution, $f(y, n)$, of the sum of n independent claims. —— Distribution, $F(y, P)$, of the sum of the claims when P claims are expected. The average claim is $50.

Example: If the expected claims under a policy are $100 ($P = 2$) and if the actual claims are $490, the probability of an experience as bad as this occurring because of chance factors is 0.1%.

3. Use summation or approximate integration to obtain $F(y, P)$ for $P = \frac{1}{2}$, 1, 2, 4, $\cdots$ by the formula

$$F(y, 2P) = \int_0^y F(x, P)F(y-x, P)\, dx.$$

4. If the calculations are carried on from both tails and if the results are plotted on probability graph paper, it is often possible to fill in the central sec-

tions by interpolation. Such interpolations should be adjusted to reproduce the correct mean. This method is illustrated in fig. 1 in the case of surgical fee insurance.

8. Summary. In this paper the Poisson distribution is generalized to allow for the assignment of varying weights to events when the number of events follows the Poisson law. The ability of the Poisson distribution to handle heterogeneous data is preserved in the generalization. An example is given showing that the distribution of certain insurance statistics agrees with that predicted by the theory.