NOTES ON MATHEMATICAL STATISTICS [1]

BY

FRANKLIN E. SATTERTHWAITE

The basis of actuarial science is statistical. For his raw material the actuary has tabulations of the behavior of certain statistics in the past. His task is to predict as accurately as possible the behavior of certain statistics in the future. If such future statistics have had identical counterparts in the past and if a large experience has been collected for these counterparts, the problem is relatively simple. The future will reproduce the past approximately. Unfortunately such an ideal condition seldom exists. Social, economic, and physical conditions change so that adjustments such as law amendment factors are necessary. Also, policy provisions and coverages change so that the statistic under study may not have existed in the past at all.

The problem is similar to that of the chemist who is asked to make nylon. First the chemist must gather together the available raw materials such as air, water, and coal. The raw materials of the actuary are the probabilities of accidents and the distributions of their costs. Next the chemist must combine his raw materials by use of complex processes which may involve intermediate products and catalysts. For the actuary the combination processes are mathematical and involve such operations as summation and integration. It is also sometimes efficient, or even necessary, to use intermediate functions such as moments and characteristic functions to obtain the result.

This paper is a collection of notes on certain mathematical techniques which have been found useful in developing, comprehending, and applying statistical theory. The specific problems taken up and the formulas developed are the same as those covered in Mr. Bailey's paper on sampling theory in this issue of the *Proceedings*. Therefore, it was considered superfluous to include an example of the actual construction of a statistical distribution. Mr. Bailey's examples are excellent and apply as well to this paper.

1. *Expected Value and Moments*: By the expected value of a statistic is meant the weighted arithmetic average value. The weights assigned to each possible value of the statistic are the probabilities that such value of the statistic will occur. Thus if a statistic $y(x)$ is a function of another statistic, x, which has a distribution function, $f(x)$, the probability that $y(x)$ will take on a particular value $y(x')$ is $f(x')$ and the expected value of $y(x)$ will be

$$E\left[y(x)\right] = \Sigma\, y(x)\, f(x), \tag{1}$$

[1] A paper on this subject was submitted in April 1941 as a thesis in lieu of parts VII and VIII of the Fellowship examinations for this Society. Because of the identity of subject matter with Mr. Bailey's paper in this volume of the Proceedings, only the technical parts of my paper are being printed at this time.

where the summation extends over all possible values of x. If x has a continuous rather than discrete distribution, the formula becomes

$$E\,[y(x)] = \int y(x)\, f(x)\, dx, \tag{2}$$

where the integration is taken over all possible values of x.

The operation symbol, E, will be reserved to indicate the expected value. If one will remember the following properties of this operator, he will find it very useful in developing many of the formulas necessary in statistical analysis. These properties are:

i The expected value operator, E, is a linear operator. That is

$$E\,(ax + b\,y) = a\,E(x) + b\,E(y). \tag{3}$$

ii The expected value of the product of two independent statistics is equal to the product of the expected values. That is

$$E(x\,y) = E(x)\,E(y). \tag{4}$$

iii The expected value of a constant is the constant itself. That is

$$E(a) = a. \tag{5}$$

iv The expected value of the expected value is the expected value. That is

$$E\,[E(x)] = E(x). \tag{6}$$

The fact that the expected value is a linear operator follows directly from the same property of integrations (or summations). We know that

$$\iint (ax + by)\, f(x,y)\, dx\, dy = a \iint x\, f(x,y)\, dx\, dy + b \iint y\, f(x,y)\, dx\, dy. \tag{7}$$

The actuary is experienced in the manipulation of linear operators through the study of finite differences. He knows that in formal manipulations he may treat the operation symbol as an algebraic quantity except that the operator and a variable may not be permuted. That is

$$\mathrm{E}(x) \neq x\,E. \tag{8}$$

The fact that the expected value of the product of independent statistics is equal to the product of the expected values follows because if two statistics are independent, the probability of their joint occurrence is the product of their probabilities of individual occurrences. Then by the separatability of integrations (or summations) we have

$$\begin{aligned} \iint x\,y\,h\,(x,y)\, dx\, dy &= \iint x\, y\, f(x)\, g(y)\, dx\, dy, \\ &= \int x\, f(x)\, dx \cdot \int y\, g(y)\, dy. \end{aligned} \tag{9}$$

Note that this property of the expected value holds only if the statistics are independent. The other properties hold for dependent as well as independent and single statistics.

The property that the expected value of a constant is equal to the constant follows from the fact that the integration (or summation) of the distribution function of a statistic over all possible values of a statistic is unity. Since the expected value is a constant, not a statistical variable, it also follows that the expected value of the expected value is the expected value.

Mr. Bailey developed his formulas for the moments of the distributions of actuarial statistics by arithmetic methods. It is very convenient to define moments and to determine relations between moments by use of the expected value notations. Thus the defining equations for the three types of moments are

$$\begin{aligned} (10) \qquad \mu''_k &= E(x-a)^k, \\ \mu'_k &= E(x)^k, \\ \mu_k &= E(x-E(x))^k. \end{aligned}$$

In this paper the Greek letter μ, (mu) will always be used to refer to moments. Unprimed it will refer to moments about the mean, with a single prime to moments about zero, and with a double prime to moments about some special origin, a. The reader should verify that the application of the properties of E gives

$$(11) \qquad \mu''_0 = \mu'_0 = \mu_0 = 1, \qquad \mu_1 = 0.$$

One should be acquainted with the following parameters and their symbols:

(12)

mean: $m = \mu'_1$,

variance = (standard deviation)^2 = (probable error/.6745)^2:

$\sigma^2 = \mu_2$,

skewness: $\beta_1 = (\alpha_3)^2 = (\mu_3)^2/\sigma^6$,

kurtosis: $\beta_2 = \alpha_4 = \mu_4/\sigma^4$.

For the normal distribution $\beta_1 = 0$ and $\beta_2 = 3$.

Facility in using the notations of expected value to obtain relationships between parameters is very useful. Note that formulas so developed are perfectly general in their application. They hold for statistics with any type of distribution function whatever. However, if the second property of E is used, the different statistics must be independent. We shall give a few examples.

$$\begin{aligned} (13) \qquad \sigma_x^2 &= \mu_2, && \text{by (12)}, \\ &= E\,[x-E(x)]^2, && \text{by (10)}, \\ &= E\,[x^2 - 2\,x\,E(x) + \{E(x)\}^2], \\ &= E(x^2) - 2\,E(x)\,E(x) + E\,\{E(x)\}^2, && \text{by (3)}, \\ &= E(x^2) - 2\,\{E(x)\}^2 + \{E(x)\}^2, \\ &= \mu'_2 - m^2, && \text{by (10)}. \end{aligned}$$

If two statistics, x and y, are independent,

(14)

$$\begin{aligned}
\sigma_{x+y}^2 &= E[(x+y)-E(x+y)]^2, && \text{by (12) and (10),}\\
&= E[\{x-E(x)\}+\{y-E(y)\}]^2, && \text{by (3) and (4),}\\
&= E[x-E(x)]^2+2\,E[x-E(x)]\cdot E[y-E(y)]+ \\
&\qquad E[y-E(y)]^2, && \text{by (3) and (6).}\\
&= \sigma_x^2+\sigma_y^2,\ \text{since } E[x-E(x)]=0,
\end{aligned}$$

The origins from which the statistics are measured are arbitrary. Therefore it will be assumed in the next two developments that they are measured from the mean so that $E(x)$ and $E(y)$ equal zero.

(15)

$$\begin{aligned}
{}_{x+y}\mu_3 &= E(x+y)^3 = E(x^3+3\,x^2y\ +3\,xy^2+y^3), && \text{by (10),}\\
&= E(x^3)+3\,E(x^2)\,E(y)+3\,E(x)\,E(y^2)+E(y^3), && \text{by (3) and (4),}\\
&= E(x^3)+E(y^3) = {}_x\mu_3+{}_y\mu_3, && \text{by (10).}\\
{}_{x+y}\mu_4 &= E(x+y)^4 = E(x^4)+0+6\,E(x^2)\,E(y^2)+0+E(y^4),\\
&= {}_x\mu_4+6\,\sigma_x^2\,\sigma_y^2+{}_y\mu_4.
\end{aligned}$$

The extension of these formulas to the distribution of the sum of n independent statistics each having the same distribution gives the following parameters where the sub-x refers to the parameters of the distribution of x:

(16)

$$\begin{aligned}
m &= n\,m_x,\\
\sigma^2 &= n\,\sigma_x^2,\\
\beta_1 &= {}_x\beta_1/n,\\
\beta_2-3 &= ({}_x\beta_2-3)/n.
\end{aligned}$$

2. *The Poisson Distribution*:

The distribution of the frequency of occurrence of independent events is of fundamental importance in the analysis of insurance statistics. This distribution is the Poisson distribution.

In order to bring out the properties of the Poisson distribution as clearly as possible, a development based on the infinitesimal calculus will be used. In following the development, the reader should remember that the values of any terms which involve infinitesimals of higher order than the first in a variable are immaterial to the argument. They drop out on integration. The

[2] The best tabulation of the Poisson distribution is: Molina, E. C.: Poisson's Exponential Binomial Limit. (1942) Van Nostrand.

notation, $o(dx)$, indicates an infinitesimal of higher order than dx and is read "zero of dx".

Let $p(\alpha)\,d\alpha$ be the probability that an event will occur in the infinitesimal unit of time, α to $\alpha + d\alpha$. Then the probability that no events will occur in this unit of time is

$$1 - p(\alpha)\,d\alpha + o(d\alpha). \tag{17}$$

Expressing this in the form of an exponential gives

$$e^{-p(\alpha)\,d\alpha\;+o(d\alpha)}. \tag{18}$$

If the probability of the occurrence of an event in any unit of time is independent of the occurrence of the event in any other unit of time, the probabilities, (18), for successive infinitesimal units of time may be multiplied together to obtain the probability that no event will occur in the finite interval of time, a to b. This is

$$\begin{aligned} F(0) &= e^{-p(a)\,da+o(da)} \cdot e^{-p(a+da)\,da\;+o(da)} \cdots\cdots e^{-p(b-da)\,da\;+o(da)}, \\ &= e^{-\int p(a)\,da\;+o(da)} = e^{-\int p(a)\,da} \\ &= e^{-P} \end{aligned} \tag{19}$$

where

$$P = \int_a^b p(\alpha)\,d\alpha \tag{20}$$

is the expected number of events.

Because of the independence conditions, $p(\alpha)\,d\alpha$ and e^{-P} may be multiplied together to obtain the probability that one event will occur in the period, α to $\alpha + d\alpha$, and that no other event will occur in the period a to b:

$$p(\alpha)\,d\alpha\, e^{-P}. \tag{21}$$

In this case α is a fixed point in time. If α is now allowed to vary over the period, a to b, the integral of (21),

$$\begin{aligned} F(1) &= \int p(\alpha)\,d\alpha\, e^{-P}, \\ &= P\,e^{-P}, \end{aligned} \tag{22}$$

gives the probability that one and only one event will occur in the period, a to b. The probability that exactly two events will occur in the period, a to b, is

$$\begin{aligned} F(2) &= \frac{1}{2!} \iint [p(\alpha')\,d\alpha']\,[p(\alpha'')\,d\alpha'']\,e^{-P}, \\ &= \frac{1}{2!}\,P^2\,e^{-P}. \end{aligned} \tag{23}$$

The $2!$ $(= 1\cdot 2)$ enters because the occurrence of the first event at time α'' and the second at time α' duplicates the occurrence of the first at time α'

and the second at α''. Similarly the probability of the occurrence of exactly n events is

$$F(n) = \frac{1}{n!} P^n e^{-P}, \quad 0! = 1, \; n! = 1.2 \ldots n. \tag{24}$$

This is the Poisson distribution. It is the distribution of the number of events *whenever* the probability of the occurence of an event is *independent* of any other occurrence of the event.

The moments of the Poisson distribution are easily determined. For example

$$\begin{aligned}
\mu'_2 = E(x^2) &= \sum_{x=0}^{\infty} x^2 \frac{P^x}{x!} e^{-P}, \\
&= e^{-P} \left\{ 0^2 + 1^2 P + \sum_{x=2}^{\infty} \frac{x + x(x-1)}{x!} P^x \right\}, \\
&= e^{-P} \left\{ P \frac{P^0}{0!} + P \sum_{x-1=1}^{\infty} \frac{P^{x-1}}{(x-1)!} + P^2 \sum_{x-2=0}^{\infty} \frac{P^{x-2}}{(x-2)!} \right\}, \\
&= e^{-P} \{ P e^P + P^2 e^P \}, \\
&= P + P^2.
\end{aligned} \tag{25}$$

The straightforward application of formulas (12) and (10) and the properties of the expected value leads to the following parameters of the Poisson distribution:

$$\begin{aligned}
m &= P, \\
\sigma^2 &= P, \\
\beta_1 &= 1/P, \\
\beta_2 - 3 &= 1/P.
\end{aligned} \tag{26}$$

Notice that no restrictions were placed on $p(\alpha)$ in this development. It can vary in any manner at all with respect to time. For any given expected number of events, P, the Poisson distribution is the same. Thus if we are studying annual accident frequency, it is immaterial to us if $p = .01$ the year around, or if $p = .02$, November to April, and $p = 0$, May to October. The assignment of α as a time variable is also arbitrary. If p (Iowa) $= .01$ and p (Massachusetts) $= .02$, the experiences may still be combined and the combined experiences will follow a Poisson distribution.

In life insurance, a person can not die twice. While he is living the probability of death is, say, $p = .01$. If he dies, it immediately changes to $p = 0$. Thus the probability of the occurrence of death depends on whether the person has died previously. If exposures are taken to the end of the year of death (as is customary), the independence conditions are not satisfied and the distribution of claim frequencies follows the binomial, not the Poisson distribution (if p is small, it is almost a Poisson distribution). If the exposure is cut off at the time of death, the independence conditions are satisfied and the distribution of the number of deaths is a Poisson distribution. This latter procedure has the added advantage that it is then perfectly proper to combine experiences with different probabilities. The combination of two experiences which follow different binomial distributions does not have a binomial distribution.

3. *Characteristic Functions*:

The characteristic function of the distribution has been found to be a very useful tool in the development of statistical theory. The characteristic function[3] is defined as

$$\phi_x(t) = E(e^{itx}), \qquad i = \sqrt{-1}. \tag{27}$$

For example, the characteristic function of the Poisson distribution is

$$\begin{aligned} \phi(t) &= \sum_{x=0}^{\infty} e^{itx} \frac{P^x}{x!} e^{-P}, \\ &= e^{-P} \sum \frac{(P e^{it})^x}{x!} = e^{-P} \cdot e^{Pe^{it}}, \\ &= e^{-P(1-e^{it})}. \qquad (28) \end{aligned}$$

The characteristic function converts the distribution function, a function of the statistic, x, into a function of a new variable, t. The characteristic function is frequently used instead of the distribution function to develop statistical theory because of two properties:

i. The characteristic function of the sum of two independent statistics is equal to the product of the characteristic functions. That is

$$\phi_{x+y}(t) = \phi_x(t) \cdot \phi_y(t), \tag{29}$$

since

$$E(e^{it(x+y)}) = E(e^{itx} \cdot e^{ity}) = E(e^{itx})\, E(e^{ity}), \qquad \text{by (4)}. \tag{30}$$

[3] If the reader has not studied functions of the complex variable and does not feel secure when manipulating "imaginary" numbers, it is suggested that he take a red pencil and cross out all the "i's" which appear in the formulas which follow. He will then have *moment generating* functions instead of characteristic functions. The reasons the mathematician has for preferring characteristic functions need not concern us here.

ii. The k^{th} moment about zero of the distribution of x is equal to $(\sqrt{-1})^{-k}$ times the k^{th} derivative of the characteristic function at $t=0$.

That is:

$$ {}_x\mu'_k = (\sqrt{-1})^{-k} \frac{d^k}{dt^k} \phi_x(t)\Big|_{t=0} \tag{31} $$

since

$$ \frac{d^k}{dt^k} \phi_x(t)\Big|_{t=0} = E\left[\frac{d^k}{dt^k} e^{itx}\right]_{t=0} = E\left[(ix)^k e^{itx}\right]_{t=0}, $$
$$ = i^k E(x^k) = i^k\, {}_x\mu'_k. $$

To illustrate the second property, let us find the first and second moments of the Poisson distribution. Differentiating (28) twice gives

$$ \frac{d}{dt}\phi(t) = e^{-P(1-e^{it})}(-P)(-1)e^{it}(i), \tag{32} $$

$$ \frac{d^2}{dt^2}\phi(t) = e^{-P(1-e^{it})}\{[(-P)(-1)e^{it}(i)]^2 + [(-P)(-1)e^{it}(i)^2]\}. $$

Setting $t=0$ and cancelling out the i's give by (31),

$$ \mu'_1 = e^{-P(1-1)}(-P)(-1)e^0 = P \tag{33} $$
$$ \mu'_2 = (1)\{[P(1)]^2 + P(1)\} = P + P^2. $$

This value of the second moment agrees with the value found in formula (25) by the direct method.

4. *The Generalized Poisson Distribution:*

In many insurance lines one is only indirectly interested in the number of events. The primary interest is centered around the total cost of the claims which arise from the events. This cost is usually a variable and is not the same for each event. The analysis of this type of situation is quite parallel to the development of the Poisson distribution given above so that the result may well be called a generalized Poisson distribution. To keep the development within bounds, it will be necessary to make use of the characteristic function (27) and its properties which were introduced in (29) and (31).

In the development of the generalized Poisson distribution the same notations and independence assumptions will be used as in the development of the Poisson distribution. Let $f(x, \alpha)$ be the distribution of the cost of a claim at time α. Then the distribution of total cost over the period, α to $\alpha + d\alpha$, is

$$F(x,\alpha)=1-p(\alpha)\,d\alpha, \qquad \text{if } x=0, \tag{34}$$
$$\qquad = p(\alpha)\,f(x,\alpha)\,d\alpha + o(d\alpha), \qquad \text{if } x>0.$$

The characteristic function of the distribution of the total claim cost over the period, α to $\alpha+d\alpha$ is by (27),

$$\Phi(t,\alpha)=(1-p(\alpha)\,d\alpha)+\int e^{itx}\,p(\alpha)\,f(x,\alpha)\,d\alpha\,dx+o(d\alpha), \tag{35}$$

since $e^{it0}=1$.
Now let

$$\phi(t,\alpha)=\int e^{itx}\,f(x,\alpha)\,dx, \tag{36}$$

so that (35) becomes

$$\Phi(t,\alpha)=1-p(\alpha)\,d\alpha\,[1-\phi(t,\alpha)]+o(d\alpha), \tag{37}$$
$$= e^{-p(\alpha)\,d\alpha\,(1-\phi(t,\alpha))+o(d\alpha)}.$$

Assuming that events in different units of time are independent, the characteristic functions, (37), for the different units may be multiplied, together in the same way as in (19), to obtain the characteristic function of the distribution of the total claim cost over the period, a to b. This gives

$$\Phi(t)=e^{-\int p(\alpha)(1-\phi(t,\alpha))\,d\alpha}, \tag{38}$$
$$= e^{-P(1-\phi(t))},$$

where

$$\begin{aligned}\phi(t) &= \int [p(\alpha)/P]\,\phi(t,\alpha)\,d\alpha,\\ &= \iint e^{itx}\,[p(\alpha)/P]\,f(x,\alpha)\,d\alpha\,dx,\\ &= \int e^{itx}\left\{\int [p(\alpha)/P]\,f(x,\alpha)\,d\alpha\right\}dx,\\ &= \int e^{itx}\,f(x)\,dx,\end{aligned} \tag{39}$$

is (assuming that the reversal of the order of integration was legitimate) the characteristic function of the mean distribution of claims,

$$f(x)=\int [p(\alpha)/P]\,f(x,\alpha)\,d\alpha. \tag{39a}$$

Thus we see that if we know the expected number of events, P, and the mean distribution of the cost of a claim, $f(x)$, the distribution (since it is determined by its characteristic function) of the total claim cost is determinate. We may combine in any way experiences with unlike distributions, $f(x,\alpha)$'s, of the cost of a claim if we can determine $f(x)$.

The application of (31) and (12) to (38) gives the following formulas for the parameters of the generalized Poisson distribution in terms of the parameters of the mean distribution of the cost of a claim. This latter distribution is referred to by a sub-x and in practice the calculation of certain intermediate parameters indicated by a sub-o will be found convenient:

(40)
$$\begin{aligned}
m &= P\, m_o, &\quad \text{where} \quad & m_o = {}_x\mu'_1,\\
\sigma^2 &= m\, \sigma_o^2, &\quad \text{where} \quad & \sigma_o^2 = {}_x\mu'_2/m_o,\\
\beta_1 &= {}_o\beta_1/m, &\quad \text{where} \quad & {}_o\beta_1 = [{}_x\mu'_3/\mathrm{m}_o]^2/\sigma_o^6,\\
\beta_2 - 3 &= ({}_o\beta_2 - 3)/m, &\quad \text{where} \quad & ({}_o\beta_2 - 3) = [{}_x\mu'_4/m_o]/\sigma_o^4.
\end{aligned}$$

For the mean, m, a single differentiation of (38) was necessary:

(41)
$$\Phi'(t) = \frac{d}{dt}\, e^{-P\,(1-\phi(t))} = e^{-P\,(1-\phi(t))}\,(-P)\,[-\phi'(t)].$$

Setting $t = 0$ and remembering that by (31)

$$\phi(t)|_{t=0} = {}_x\mu'_0 = 1,\ \ \phi'(t)|_{t=0} = i\, {}_x\mu'_1,$$

gives the above formula. A second differentiation of (41) gives

(41a)
$$\Phi''(t) = e^{-P\,(1-\phi(t))}\,\{(-P)^2\,[-\phi'(t)]^2 + (-P)\,[-\phi''(t)]\},$$

which on the application of (31) gives

$$i^2\, \mu'_2 = i^2\,(P{}_x\mu'_1)^2 + i^2\, P{}_x\mu'_2.$$

Cancelling the i^2's and applying (12) gives the above formula for σ^2. Similarly the above formulas for β_1 and β_2 can be obtained.

The above development of the generalized Poisson shows its properties but is more or less useless for purposes of calculation. Therefore consider the repeated application of formula (29). This gives the characteristic function of the distribution of the sum of n independent statistics from the same population as ϕ^n where ϕ is the characteristic function of the distribution of a single claim. Now if n is considered a statistical variable which obeys the Poisson distribution, (24), the expected value of ϕ^n with respect to (24) is

(42)
$$\begin{aligned}
\Phi &= E(\phi^n) = \Sigma\, \frac{1}{n!}\,(P\,\phi)^n\, e^{-P},\\
&= e^{P\phi}\cdot e^{-P},\\
&= e^{-P(1-\phi)},
\end{aligned}$$

which is identical with the characteristic function of the generalized Poisson given in (38). Thus, to calculate the numerical values of a generalized Poisson distribution:

i. Find the distribution of the sum of zero, one, two, three, . . ., claims.
ii. Multiply each of these by the corresponding terms of (24).
iii. Add together these products.

Also consider the characteristic function of the distribution of the sum of two independent statistics, each of which obeys a generalized Poisson distribution with the characteristic function (38). By the application of (29) this is

$$\Phi = [e^{-P(1-\phi)}]^2 = e^{-2P(1-\phi)}. \tag{43}$$

This is identical with the characteristic function of the generalized Poisson distribution when $2P$ claims are expected instead of P claims. Therefore in calculating the generalized Poisson distribution for $2P$ expected claims, it may be considered as the distribution of the sum of two independent statistics obeying generalized Poisson distributions with P claims expected.

The distributions of the total claim payments in practically all lines of insurance fall in the class of generalized Poisson distributions. They are such directly if each claim payment is independent of the rest. If certain payments are related to each other, they can be combined and the distribution of their sum taken as the element. Thus a group health and accident policy may provide for weekly indemnity, hospital payments, and surgical payments. For a given sickness or accident these three payments are related. However if one first finds the distribution of the total of the three payments for each sickness or accident, this distribution can be used as the basis of a generalized Poisson distribution.

5. *The Hyper-geometric Distribution*:

The Poisson distribution is a one parameter distribution. If the expected number of events, P, is known, one can completely specify the distribution. However in many practical problems one does not know the expected number of events, but only knows an estimate of it. For example, in group life insurance the expected number of deaths under a policy can be calculated in accordance with some general mortality experience. However, this is only an estimate of the true expected number of deaths in the sense of theoretical statistics. It does not make allowance for the individual characteristics of the risk such as its geographical location, the type of personnel hired and working conditions. Allowance can be made for these by assuming that P is a statistical variable distributed about the estimate, N, as a mean. Unless we have a great deal of information available, we can not determine much about the distribution of P. From general reasoning we know it is non-negative and continuous. One of the simplest such distributions is the Type III distribution,

$$g(P)dP = \frac{1}{(b-1)!}\left(\frac{b}{N}\right)^b P^{b-1} e^{-bP/N}\,dP, \tag{44}$$

which has a variance (in units of the mean) of

$$\sigma^2/m^2 = 1/b. \tag{45}$$

The expected value of the Poisson distribution, (24), with respect to (44) is

$$F(n) = E\left\{\frac{1}{n!} P^n e^{-P}\right\} = \frac{(n+b-1)!}{n!\,(b-1)!}\left(\frac{b}{N}\right)^b\left(\frac{N}{b+N}\right)^{b+n} \tag{46}$$

which is called the hyper-geometric distribution. It has the parameters:

$$\begin{aligned} m &= N, \\ \sigma^2/m^2 &= 1/N + 1/b, \\ \mu_3 &= N + 3\,\frac{N^2}{b} + 2\,\frac{N^3}{b^2}. \end{aligned} \tag{47}$$

Similar treatment of the characteristic function of the generalized Poisson distribution, (38), gives the characteristic function of the generalized hyper-geometric distribution,

$$\begin{aligned} \Phi(t) &= E(e^{P(1-\phi)}) \\ &= \left[1 + \frac{N(1-\phi)}{b}\right]^{-b} \end{aligned} \tag{48}$$

The application of (31) and (12) to this gives the parameters:

$$\begin{aligned} m &= N\,m_x \\ \sigma^2 &= m\,\sigma_o^2 + m^2/b, \qquad\qquad \sigma_o^2 = {}_x\mu'_2/{}_xm \\ \mu_3 &= N\,{}_x\mu'_3 + 3\,\frac{N^2}{b}\,{}_x\mu_2'\,m_x + 2\,\frac{N^3}{b^2}\,m_x^3. \end{aligned} \tag{49}$$

In practice the parameter, b, which appears in the above formulas can be estimated in the following way:

1. Tabulate the actual value, y_j, (viz., total claims under a policy) and the expected value, m_j, (viz., pure premiums) of a number of statistics which follow the desired distribution. It is best that the m_j's should not vary too much.
2. Calculate
$$R^2 = \Sigma\,(y_j - m_j)^2,\ S^2 = \Sigma\,m_j^2,\ T = \Sigma\,m_j.$$
3. The expected value of R^2 is

$$\begin{aligned} E(R)^2 &= \mathrm{E}\,[\Sigma\,(y_j^2 - 2\,y_j\,m_j + m_j^2)], \\ &= \Sigma\,[E(y_j^2) - 2\,E(y_j)\,m_j + m_j^2], && \text{by (3)}, \\ &= \Sigma\,[m_j\,\sigma_o^2 + \frac{m_j^2}{b} + m_j^2 - 2\,m_j^2 + m_j^2], && \text{see (49) and (13)}, \\ &= T\,\sigma_o^2 + S^2/b. \end{aligned} \tag{50}$$

4. Set R^2 equal to its expected value and solve for b.

When b is large compared with the expected number of claims under each policy, statistical methods for the determination of b breakdown because the difference, $R^2 - T\sigma_o^2$, is small compared with the variability of R^2. It is then necessary to fall back on personal judgment in the choice of b. For example, if it is estimated that the true measure of the risk for 95% of the policyholders is within 20% of the rate on which manual premiums are based, the standard deviation, σ, (in units of the mean) of the distributions of the true risks about the manual rate is approximately 10%. Then, applying formula (45), we see that b is approximately 100.

6. *Summary*

The practical problem of the actuary is to forecast the behavior of certain statistics in the future. These statistics can usually be expressed as functions of certain elementary statistics. Fortunately the volume of past experience available for these elementary statistics is usually quite large so that they may be studied in great detail. Using the distributions of these elementary statistics as a basis, it is then possible to do the combining and integrating necessary to find the distribution of the desired statistic by mathematical formula or by mean strength on the calculating machine. The result is obtained more quickly (and often more accurately) than would be possible by waiting for a sizable experience to accumulate for the particular statistic under study.

The distribution of the frequency of events was found to take a very simple and general form, the Poisson distribution. The only information necessary to completely specify this distribution is the expected number of events. If the expected number of events is unknown but can be estimated, the hyper-geometric distribution is used in place of the Poisson. Both the Poisson and hyper-geometric distributions generalize to give the distribution of the total of claim payments. These generalized distributions were found to be functions of the average distribution of claims alone so that questions of seasonal or other variations in the distribution of claims can be ignored.

In some circles there has been a tendency to disparage the actuary as being backward in adapting theoretical methods to his problems. However, those who study the problem more carefully discover that the inadequacy is as much with statistical theory than with the actuary. Mathematical statistics is an infant science which has only reached the stage of rapid development in the last ten or twenty years. It still has many very simple types of problems to solve. It is only beginning to develop methods which are general enough to handle the complex problems to which the actuary must obtain an answer the best he can, inadequate theory being no excuse.

It is hoped that the reader will have been encouraged to reflect on the theoretical basis of his work more often and that maybe he will take some part in expanding that theoretical basis.