INDIVIDUAL RISK THEORY AND RISK STATISTICS AS APPLIED TO FIRE INSURANCE

BERTIL ALMER

Stockholm, Sweden

INTRODUCTION

Risk Theory for Life Insurance is simplified by the fact that the distribution $\Psi(x)$ of claim amounts x approximately coincides with the distribution of "Risk sums" (not exactly, owing to differaences in the claim frequency with age and actual state of health), ond this distribution is comparatively stable *).—The dependence on the claim frequency is eliminated by the introduction of a new time variable, and the system reduced to a (stationary) Poisson Process, which should be valid at least for *large* risk systems and for the total Life branch for a moderate sequence of years.

In almost all non-life branches, partial claims will dominate and $\Psi(x)$ can only be determined *by risk statistics,* leaving a certain *space of indetermination,* in particular for *large claims* and for mediumsized *statistical risk groups.*

In my *previous analyses,* in particular New York 1957, interest has been concentrated on *traffic and motor car insurance,* where the risk depends on cars insured *and* on the meeting traffic (including road conditions). In one year the same car can be involved in many accidents and double claims (=collisions) are rather frequent.— According to my experience, this system is best represented by a sequence of single and double *risk situations in time* (for individual cars *or* for risk groups).

Analysis is simplified for *Fire Insurance (and many other non-life branches),* because the risk system is composed of *mostly independent insurances (or risk objects),* which are best described by the ordinary *Individual Risk Theory.*

In the following text, I will try to give methods and formulas, which are the result from many years of *practical risk analysis* and

---

*) I have conserved the notation $\Psi(x)$ from New York 1957 in order to simplify eventual comparisons.

studies in risk statistics. In particular, I will try to prove that *numerical analysis in Individual Risk Theory* is almost as simple as in *Collective Risk Theory*, although the risk system is non-stationary and non-homogeneous.

THE RISK MODEL

Individual Risk Theory analyses *a finite system of Risk Objects,* representing insurances or part of insurances.

Every *risk object is characterized* by some *parameters*, representing *tariff classes* and well-defined *statistical parameters*. In analysis, however, continuous parameters are always reduced to risk intervals, owing to the fact that *risk statistics* are always limited to a fairly small number of *risk groups* or risk classes. Further subdivision will lead to *diffuse figures*, where the stochastic error will generally exceed figures obtained by statistics.

Claims are generated from these *risk objects* under the following *Hypotheses*: (I) *For every fixed period—in particular for every insurance year—and for every risk object (i) there exists a well-defined probability* (ρ_i) *for the occurrence of one claim and in case of a claim occurred also for the distribution* $\Psi_i(x)$ *of the claim amount* (x).

(2) *All claims in fire (and other non-life branches) are positive* (x $>$ o) and uniformly bounded.

(3) *The corresponding probability for two, eventually three, claims* $(\rho_{i2}\ \rho_{i3}\ \Psi_{i2}(x)\ \Psi_{i3}(x)$ *is small in comparison with* ρ_i *and is often neglected.*

(4) *Contagion from neighbouring risks is included in* ρ_i—*if this risk is insured in the same risk system there will be a small error or correction for number of claims but the claim amount is given correctly. Otherwise risk objects are independent.*

From this system follows that individual risk theory is constructed with a minimum of hypotheses, in principle reduced to the *existence* of a definite probability for claims and claim amounts *for each individual risk object, and for every statistic period.*

This *existence* is proved or at least made extremely plausible, if we start from the modern concepts of physical science and neglect the problem of free will. In this case the probabilities will exist

as definite probabilities at the beginning of each insurance year. On the other hand, the risk for each individual risk object for a sequence of time intervals will oscillate and evolve with an unknown trend. Evidently the system is well defined, but not stationary, and probably the risk in fifty years will be radically changed.

Even if we admit free will, most arbitrary changes will occur at definite times *before* the corresponding *risk situations*, for example choice of new tenants in a house insured. In that case definite probabilities will still exist before claims, although not at the beginning of the insurance year.

Finally, remaining indetermination might be expressed by an error term Q', similar to the term Q which is analysed in a following section as an effect of inhomogenity.

COMPARISON WITH RISK STATISTICS

In most cases risk statistics are analysed by punched cards, with one card for *each period and risk object* insured, and one card for every claim occurred. These cards are divided into a series of risk groups by tariff classes and eventually some additional risk parameters, and *the statistic material* takes the form

Risk classes and year $= \alpha\,\beta\,\gamma\,\delta$ (=indices)

Risk objects within a risk class, number $= N$

total sum insured $= S$

Claims occurred d: o   number $= n^*$

individual amounts $= x_j{}^*$ $(j = 1,\ldots n^*)$

total amount $= y^* = \Sigma\,x_j{}^*$,,   ,,

approximate distribution $= \overline{\Psi}{}^*\,(x)$.

Derived quotients: loss frequency $= \varphi^* = n^*/N$ in ${}^0/_{00}$

mean value $= \bar{x}^* = y^*/n^*$

risk premium $= \gamma^* = y^*/N$

d: o relative $= q^* = y^*/S$ in ${}^0/_{00}$

where * signifies statistical values $=$ *values a posteriori*.

For an independent system of risk objects (within the risk group $(\alpha\,\beta\,\gamma\,\delta)$ the corresponding mean *values a priori* are expressed by

$$n = \Sigma\,\rho_i \quad n\cdot\overline{\Psi}\,(x) = \Sigma\,\rho_i\,\Psi_i(x) \quad {}_i = 1,2,\ldots N\alpha\beta\gamma\delta$$
$$y = \Sigma\,\rho_i\cdot\bar{x}_i \text{ where } \bar{x}_i = \int x\,d\,\Psi_i\,(x)$$

Observe that $\overline{\Psi}(x)$ represents the mean conditional distribution of *one* claim once a claim has occurred. The distributions of n and y are obtained by more complicated methods, $Cf\ \theta\ (y)$.

Inversely figures from risk statistics represent approximations of risk figures which can be used as a basis for risk analysis.

This fact explains, why analysts with extensive practical experience, always accentuate *the mean values or a priori values*

$$n,\ \overline{x},\ \mathrm{y} = n\ .\ \overline{x},\ \overline{\Psi}(x)\ \text{and}\ n^*,\ \overline{x}^*,\ \mathrm{y}^* = n^*\ .\ \overline{x}^*,\ \overline{\Psi}^*(x)$$

as the *principal object for risk statistics,* and also for an independent system that, when joining arbitrary risk groups,

$$N,\ S,\ -\ n,\ \mathrm{y} = n\overline{x},\ n\ .\ \overline{\Psi}(x),\ -\ n^*,\ \mathrm{y}^* = n^*\ .\ \overline{x}^*,\ n^*.\overline{\Psi}^*(x)$$

are additive.

By this comparison we have defined the principal object of risk theory as

a) the determination of the distribution of total claim amount $\theta\ (\mathrm{y})$ from n and $\Psi\ (x)$.

b) aids in the analysis of risk structure from statistical figures for *a system* $(\alpha\ \beta\ \gamma\ \delta)$ of *risk groups* or *risk parameters.*

QUESTION OF INHOMOGENITY

Risk theory is simplified if we assume that statistical risk classes are homogeneous, and this hypothesis is reasonable in preliminary analysis. *A direct analysis* of homogenity, however, is possible in branches with a large claim frequency, say, 100 $^0/_{00}$ per annum or more, in particular motor car and accident insurance—simply by comparing the relative frequency of *risk objects with no,* 1, 2, 3, *etc. claims* in the same period (year) with the common *Poisson distribution*. Different figures have been published from Sweden, Germany, USA and other countries, proving an inhomogenity, more or less similar to my own *figures for motor car*

75 % with frequency $= \varphi_0$, 24,5 % with $3.\varphi_0$, 0,5 % with $\varphi > 1$.

In some analyses the influence of the last group (=very bad risks) covers an important part of the total risk volume (=claim amount).

In fire insurance where the claim frequency for a risk group varies from 1 $^0/_{00}$ to, say, 10 $^0/_{00}$, a direct analysis seems impossible.

On the other hand, logical analysis of a well-defined risk group—
say one-family houses—makes it extremely possible that the risk
will vary with small differences in construction and still more with
inhabitants and their manner of living. Assuming a large group
of 10.000 similar houses with the claim frequency $1 \, {}^0/_{00}$ and an
(unknown) small group of 100 houses of very bad risks with the
frequency $100 \, {}^0/_{00}$, both will lead to an average number of 10
claims a year. Evidently stochastic variation of number and amount
of claims is strongly influenced by a small number of bad risks.

Probably still more common is the system, where most risk
objects are much better than the mean risk = 1 to $2 \, {}^0/_{00}$, and this
frequency generated by a fairly small number of second rate risks.

From these examples it is easy to derive a *measure of inhomogenity*.
With the same notations as above we have for the analysed *risk
group of N objects*

$$\sum_{i=1}^{N} \rho_i = n \quad \frac{n^2}{N} \leq \sum_{i=1}^{N} \rho_i^2 \leq n^2 \quad (0 < \rho_i < 1)$$

and putting $\sum \rho_i^2 = n^2.Q \leq n^2/L$ *(L = integer) Q represents an
exact and L an approximate measure of inhomogenity*. At the same
time $\sum_{j>i} \rho_i.\rho_j = \frac{1}{2} (n^2 - \sum \rho_i^2)$.

As a quadratic measure Q has some properties in common with
the variance, although the decrease with growing material is less
regular, owing to the influence of very bad risks.

We might also observe that a priori risk values should be regarded
as (unknown) *constants*, and consequently the maximum and mini-
mum of the variation under the condition Q (or L) is well-defined
and possible to calculate—contrary to σ^2 which only gives limits
for mean values and no information on exact variation for some
few "stochastic experiments".

Individual risk theory—reduction to standard distributions:

An independent system of risk objects can be analysed by for-
mulas similar to generating functions.

The combined risk for *number of claims n* is expressed by the
following formulas, where the *coefficient of* λ^k expresses the pro-
bability of exactly k claims.

$$\prod_{i=1}^{N} \left((\mathrm{I} - \rho_i) + \lambda \rho_i \right) = \prod_{i=1}^{N} \frac{\mathrm{I} + \lambda q_i}{\mathrm{I} + q_i} \text{ where } q_i = \frac{\rho_i}{\mathrm{I} - \rho_i}$$

$$= A_0 + A_1\lambda + A_2\lambda^2 + \dots A_N\lambda^N = A_0 (\mathrm{I} + B_1\lambda + B_2\lambda^2 + \dots B_N\lambda^N)$$

(An interesting fact is that these formulas can be derived by elementary probability theory).

The successive coefficients B_k are symmetric functions of ρ_i and all ρ_i are positive.

In previous papers (New York 1957) I have analysed the *extrema of symmetric functions of positive variables* under the conditions

$$\sum_{i=1}^{N} q_i = n'^1; \; \sum_{i=1}^{N} q_i^2 = n'^2 \;.\; Q \leq n'^2/L' \; (L' = \text{integer})$$

and proved that *maximum* is attained for the system with *maximum symmetry*

$$q'_i = n'/N \; (i = 1,2,3,\dots N)$$

and *minimum* is obtained for the system with *maximum asymmetry*

$$q'_i = n'/L \text{ for } i \leq L$$
$$q'_i = O \quad \text{ for } i > L$$

By this theorem *further analysis is reduced to two homogeneous binomial systems* with the claim frequency

$$q^{(1)} = n'/N \text{ and } q^{(2)} = n'/L$$

respectively. For a large risk group (N large) the upper limit is approximately equal to the ordinary *exponential or Poisson distribution*.

The approximation error depends on the lower limit which depends on the measure of homogenity $(=L)$, *or on the estimate of the degree of inhomogenity*. For *the total system (or total fire branch)* utilized in *ruin problems*—in particular the question of an addition to premiums for safety—L might be estimated from the risk frequency for different risk groups with some estimated additions.

The corresponding analysis of the distribution $\theta(y)$ *for the total sum of claims* starts from the analogeous formula

$$\prod_{i=1}^{N} ((\mathrm{I} - \rho_i) + \lambda \, \rho_i{}^*) \, \Psi_i (x)$$

where, however, products of distribution functions, e.g.

$$\Psi_i\,(x)^* \;\Psi_j(x) \;=\; \int_{-\infty}^{+\infty} \Psi_i\,(x\text{-}t)\;d\;\Psi_j\;(t) \;=\; \int_{0}^{x} \Psi_i\,(x\text{-}t)\;d\;\Psi_j\;(t)$$

should be calculated by *convolutions=asterisk products*,—or eventually $\Psi_i\,(x)$ are replaced by the corresponding characteristic functions.

The extremum theorem is still valid if we substitute

$$\Sigma\,\rho_i\,\Psi_i\,(x) \;=\; n\,(x) \;=\; n\,.\,\overline{\Psi}\,(x) \qquad \text{for } \Sigma\,\rho_i = n$$

and $\Sigma\,\rho_i{}^2\,\Psi_i\,(x)^*\,\Psi_i\,(x) \;=\; [n\,(x)^*\,n\,(x)]\,.\,Q\,(x)$

$$\leqslant n\,(x)^*\,n\,(x)\,/\,L' \;=\; n^2\,.\,\overline{\Psi}\,(x)^*\,\overline{\Psi}\,(x)/L \quad \text{for } \Sigma\,\rho_i{}^2 < n^2 Q \leq n'^2/L'$$

leading to the same results. The binomial distributions take the form

$$\{(\mathrm{I} - \rho) + \lambda\,\rho^*\,\overrightarrow{\Psi}\,(x)\}^k \text{ with } \rho = \frac{n}{K}\,;\; \overline{\Psi}\,(x) = \frac{\Sigma\,\rho_i\,\overline{\Psi_i}\,(x)}{n}$$

$$\text{and } K = N \text{ or } K = L.$$

In the further analysis *the upper limit will lead to Bessel functions* and *the lower limit to degenerate Bessel functions*.

Individual risk theory—approximation of $\Psi\,(x)$

In this section analysis is *restricted to homogeneous binomial distributions*—or more restrictively to the limit distributions = *exponential or Poisson distributions*.

Starting from the well-known formulas

$$\frac{x^m}{m!} * \frac{x^n}{n!} = \int_{0}^{x} \frac{t^m\,(x-t)^n}{m!\;n!}\,dt = \frac{x^{m+n+1}}{(m+n+\mathrm{I})!}$$

$$\left(\frac{x^m}{m!}\cdot e^{-x}\right) * \left(\frac{x^n}{n!}\cdot e^{-x}\right) = \int_{0}^{x} \frac{t^m\,(x-t)^n}{m!\;n!}\,e^{-t-(x-t)}\,dt = e^{-x}\int_{0}^{x} \frac{t^m\,(x-t)^n}{m!\;n!}\,dt$$

it seems well-founded to call these functions "fundamental functions for the operation of asterisk multiplication or convolution". In some researches the slightly more complicated function $\dfrac{x^m}{m!}\,\dfrac{(z-x)^n}{n!}$ is also utilized.

In the following researches $\overline{\Psi}$ (x) is *normalized* in the inverse sense, so that $\Psi(\infty) = 0$ and $\Psi(0) = 1$

The above observation makes it extremely probable that *calculations* should be *simplified if claims distributions* $\overline{\Psi}$ (x) *are approximated* (*and equalized*) *by exponential polynomials* (including the degenerated form $x^m(z\text{-}x)^n$), and in my *researches from* 1950 I have made extensive tests of this form, and also as an "extremal case"
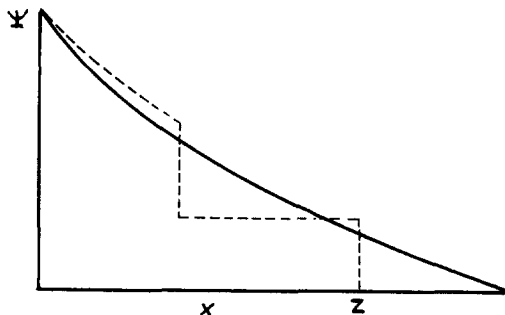
of the *Pareto distribution*. Later I have abandoned Pareto, except in the study of monetary changes, partly because for a system of finite claims $(x \leq Z)$ well approximating parts are separated by steep drops, and partly because this "step curve" will give about the same results as an equalized exponential curve, extended to infinity (all insurance claims are uniformly bounded).

My own researches, extended to accident, sickness, third party, "traffic" and motor car insurance seem to prove that except for exceptional branches,

all distributions $\overline{\Psi}$ (x) determined from an extensive statistic material *can be approximated and equalized by an exponential trinomial*

$$\overline{\Psi}(x) = nA_1 \cdot \beta_1\, e^{-\beta_1 x} + nA_2 \cdot \beta_2\, e^{-\beta_2 x} + n\, A_3 \cdot \beta_3\, e^{-\beta_3 x}$$

$$A_1 + A_2 + A_3 = 1$$

with a relative error of some percent.

These words *relative error* are all-important for applications (for the infinite branch "relative error" is exchanged to "relative total error, as compared with claims amount for moderate intervals in x utilized as subdivisions").

In some cases the last term should include a factor x^k.

Factor A_1, A_2, A_3 are of the same order of magnitude and $n \cdot A_1$, $n \cdot A_2$, $n \cdot A_3$ can be considered as *partial numbers of claims.*

The other principal *form of exponential polynomials*

$$(A_1 x^k + A_2 x^l + A_3 x^m) e^{-x}$$

is less analysed but *when applicable will lead to slightly simplified calculations.*

A practical example was given in my New York paper 1957 in the form (traffic insurance).

$$7{,}86 \cdot \overline{\Psi} (x) = 2.83 \cdot 0.046 \cdot e^{-0.046x} + 1.46 \cdot 0.35 \cdot e^{-0.35x} + 1.73 \cdot$$
$$\cdot \, 1.04 \cdot e^{-1.04x} + 1.84 \cdot 3.31 \cdot e^{-3.91x} \text{ (maximum difference } = \text{1-2 \%).}$$

We might accentuate that coefficients in the exponents are well separated (by a factor > 3). Consequently, the system is easily calculated by successive approximation—the two first terms from large claims, the last term from small claims while the third term is very sensitive to correct *adaptation.*—Also the separate terms characterize different intervals of x, which even can be identified as different types of claims (*Cf.* New York 1957).

INDIVIDUAL RISK THEORY ∞ COLLECTIVE RISK THEORY—*transformation formulas for the system* $[n, \overline{\Psi}(x)] \to \theta \, (x)]$

The Poisson distribution is univocly characterized by either of the following properties:

a) When two Poisson systems for arbitrary material or periods

$$\left[n_1, \overline{\Psi}_1 (x) \right] + \left[n_2, \Psi_2 (x) \right] = \left[(n_1 + n_2), \frac{n_1 \Psi_1 (x) + n_2 \Psi_2 (x)}{n_1 + n_2} \right]$$

are joined, the two characteristic values n = mean value of number of claims (= a priori number) and $n \cdot \overline{\Psi} (x)$ = mean distribution of claim amount for a single claim multiplied by n, are obtained by addition. $y = n \cdot \overline{x}$ is also additive.

b) A Poisson system for an arbitrary period can be calculated by a time-homogeneous system (based on the mean values n and $\overline{\Psi} \, (x)$ for this fixed period — although *both* change from year to year).

Assuming that one individual risk system has converged to the limit (or Poisson) system (except for a determined error term), the transformation from mean values a priori n and $\overline{\Psi}(x) \sim$ approximate statistical figures—to the distributions of n and y (=total claim amount=sum of claims) is expressed by the well-known series:

$$\theta(y) = e^{-n} \cdot \sum_{k=0}^{\infty} \frac{n^k}{k!} [\overline{\Psi}(x)]^{k*} \text{ where } [\overline{\Psi}(x)]^{k*} = [\overline{\Psi}(x)]^{(k-1)*} *\Psi(x)$$

where the first factor $e^{-n} \cdot \dfrac{n^k}{k!}$ expresses the probability of exactly k claims and $[\overline{\Psi}(x)]^{k*}$ the corresponding distribution of claim amount $(y = x)$.

Remark: This formula, combined with the following method of calculation, represents my contribution to the excess problem with exceptional number of claims or exceptional claims values—only I find it very difficult to combine very small probabilities with any reasonable explanation of practical consequences—probably values should be interpreted as measures, which might be utilized in the comparison of different re-insurance systems.

The transformation problem is easily resolved for the fundamental functions, for example $[n, \beta e^{-\beta x}]$ is resolved by

$$\beta \cdot e^{-n-\beta x} \cdot \sum_{k=1}^{\infty} \frac{n^k}{k!} \cdot \frac{(\beta x)^{k-1}}{(k-1)!} = \beta n \cdot e^{-n-\beta x} \sum_{k=0}^{\infty} \frac{(\beta n x)^k}{k!\,(k+1)!} =$$

$$= \theta(n, \beta, x) = \beta n \cdot e^{-n-\beta x} \cdot I_0(2\sqrt{\beta n x})$$

where $I_0(2\sqrt{\beta n x})$ = imaginary Bessel function of zero order. *Different forms of $\theta(y)$ are easily derived from this expression, also the important distribution of claim mean value $(= \overline{x})$.*

In order to extend the transformation to *exponential trinomials* we should observe that the Poisson system

$$[n, \overline{\Psi}(x) = A_1 n \cdot \beta_1 e^{-\beta_1 x} + A_2 n \cdot \beta_2 e^{-\beta_2 x} + A_3 n \cdot \beta_3 e^{-\beta_3 x}]$$

is *generated by joining three fundamental systems*

$$\theta(A_1 n; \beta_1; x) \quad \theta(A_2 n; \beta_2; x) \quad \theta(A_3 n; \beta_3; x)$$

proving that *the joint distribution is expressed by*

$$\theta\,(A_1\,n;\,\beta_1;\,x)\,{}_*\,\theta\,(A_2\,n;\,\beta_2;\,x)\,{}_*\,\theta\,(A_3\,n;\,\beta_3;\,x)$$

This asterisk product is easily calculated for a series of values n by Hollerith multiplier or by electronic calculators, and the program has a form that will express the joint distribution θ (x) as a function of *x*. The *complete table* also contains a series of values of *n*. (This table applied to Collective Risk Theory gives a solution for moderate *and* large time intervals).

By this system *I hope to have presented a (first) complete numerical solution of the transformation problem*

$$[n,\,\overline{\Psi}\,(x)] \to \theta\,(y)$$

applicable uniformly for small and large values of n, for approximate or exact Poisson distributions. (First communicated NTA=to some Scandinavian actuaries, in 1956).

(I have always been interested in this problem because for me it has seemed evident that the ordinary asymptotic formulas by Esscher and others will always give a good approximation for a *very large* system in the neighbourhood of the mean value, say for $m \pm \sigma$ or even $m \pm 2\sigma$ — and also, say, outside $m \pm 6\sigma$ (=0). Most applications, however, apply to small probabilities (0,01 to 0,0001) and in this space homogenizing is less advanced, and a small absolute error is not sufficient to prove a small relative error.— The question is easily *tested* by above exact methods).

Some hints for numerical analysis

(1) As a first approximation of $\overline{\Psi}\,(x)$* we might use two exponential monomials or fundamental functions, representing an upper limit $\Psi^+\,(x)$ and a lower limit $\Psi^-\,(x)$. The final θ (y) must lie between $\theta^+\,(y)$ and $\theta^-\,(y)$. The resulting probabilities for differences in $\overline{y}$ from the mean value *y* exceeding a fixed value, are often obtained with quite acceptable precision.

(2) In most researches *the statistic figures will give an approximation of* $\Psi\,(x)$, which is more or less exact for small and medium sized x, but uncertain for large *x*. In this case, we should introduce either a variable factor (1 $\pm$ ε) in the first term of the trinomial or else an extra error term

$$A\,.e^{-\beta\,.x/2}\quad(=\text{square root error}).$$

Representing the *estimated upper and lower limits* of the empirical distribution curve by $\Psi^+(x)$ and $\Psi^-(x)$, a corresponding error term is obtained for $\theta(y)$ and for any calculated "probability that the error will not exceed a given limit".

(3) In particular *if the first form of error term is utilized*, the above-mentioned error can also be expressed by a modified value of y equivalent to an error

$$n\,(\text{I} \pm \delta_1) \qquad x\,(\text{I} \pm \delta) \qquad \delta_1 + \delta_2 \backsim \varepsilon\,(x)$$

The somewhat surprising result is, that *errors in the empirical values* $n, \overline{x}$ *and* $\overline{\Psi}(x)$ *will enter in the final expressions in much the same way*—very comforting for the poor actuarial conscience

AIDS FOR STRUCTURE ANALYSIS, IN PARTICULAR FACTOR METHODS

American tariffs are often constructed in the factor form (e.g. motor car and fire):

Starting from a short *series of basic tariffs*—calculated from *the total statistics for these tariff classes,—modifications in per cent* are introduced for quite a number of secondary "risk factors" or parameters, and these in their turn are calculated from the joint material of insurances, including the specified parameter. The actual tariff premium equals this product for all actual parameters.

Probably, this *"factor system"* starts from *the concept of mutually independent risk factors*. It is *not evident*, however, that this system *constitutes an equitable tariff system*, although American risk material should be sufficiently large to *furnish equalized results* for tariff companies.

My analyses, however, for traffic and motor car in 1950, seemed to indicate some factor dependence, and were *developed to a method for structure analysis in factor form*, including numerical calculation.

Assuming, that risk statistics are produced by Hollerith cards in the form (for notations *vide* previous section):

$$N_{\alpha\beta\gamma\delta} \quad n^*_{\alpha\beta\gamma\delta} \quad y^*_{\alpha\beta\gamma\delta} \quad \text{absolute and additive}$$

$$\varphi^*_{\alpha\beta\gamma\delta} \quad \overline{x}^*_{\alpha\beta\gamma\delta} \quad \gamma^*_{\alpha\beta\gamma\delta} \quad \text{relative, non-additive}$$

where eventually α = tariff, β = geographical district, γ = car value,

δ=year, and that all parameters involve from 3 to 10 alternatives, we will indicate sums according to one or more parameters by excluding the corresponding indices.

We will now *introduce the factor hypothesis* in the form

$$\varphi^*_{\alpha\beta\gamma\delta} \curvearrowright \tilde{\varphi}_{\alpha\beta\gamma\delta} \equiv \tilde{\varphi} \cdot A_\alpha \cdot B_\beta \cdot C_\gamma \cdot D_\delta \quad \tilde{\varphi} = n^*/N$$

$$\bar{x}^*_{\alpha\beta\gamma\delta} \curvearrowright \tilde{x}_{\alpha\beta\gamma\delta} \equiv \tilde{x} \cdot A'_\alpha \cdot B'_\beta \cdot C'_\gamma \cdot D'_\delta \quad \tilde{x} = y^*/n^*$$

and eventually

$$\gamma^*_{\alpha\beta\lambda\delta} \curvearrowright \tilde{\gamma}_{\alpha\beta\gamma\delta} \equiv \tilde{\gamma} \cdot A''_\alpha \cdot B''_\beta \cdot C''_\gamma \cdot D''_\delta \quad \tilde{\gamma} = y^*/N$$

Calculations are now performed by *successive approximations,* or as I prefer to say *"par la méthode de balayage de M. Poincaré."* Loss frequency and mean value (or risk premium) are here considered as a field of crossing forces in the field (α, β, γ, δ), emanating from the four margins in α β γ δ respectively.

As first approximation we might take the American margin system, identifying A_α B_β C_γ D_δ etc with their margin values. The obtained approximation, however, is not very good for the simple reason that the basic distribution of $N_{\alpha\beta\gamma\delta}$ is asymmetric and non-stochastic. (Values might be improved by the following method)

The basis for successive approximation is the previous observation that $n^*_{\alpha\beta\gamma\delta}$ and $y^*_{\alpha\beta\gamma\delta}$ are the only *additive* values. Consequently, all calculations are performed not on relative but with absolute values=*any intermediate factor table* (i) is utilized to calculate the resulting *approximations in n and y.*

$$_i\tilde{n}_{\alpha\beta\gamma\delta} = N_{\alpha\beta\gamma\delta} \cdot \tilde{\varphi} \cdot {_iA_x} \cdot {_iB_\beta} \cdot {_iC_\gamma} \cdot {_iD_\delta} \text{ compared with } n^*_{\alpha\beta\gamma\delta}$$

$$_i\bar{y}_{\alpha\beta\gamma\delta} = n^*_{\alpha\beta\gamma\delta} \cdot \tilde{x} \cdot {_iA'_x} \cdot {_iB'_\beta} \cdot {_iC'_\gamma} \cdot {_iD'_\delta} \text{ compared with } y^*_{\alpha\beta\gamma\delta}$$

The resulting approximation, say in α, is tested by

$$_i\tilde{n}_\alpha \curvearrowright n^*_\alpha \quad (\alpha = 1,2,3,....)$$

$$_i\tilde{y}_\alpha \curvearrowright y^*_\alpha \quad (\alpha = 1,2,3,....)$$

These sums furnish *improved factors in* α

$$_{i+1}A_\alpha \text{ ledaing to } _{i+1}\tilde{n}_{\alpha\beta\gamma\delta} = N_{\alpha\beta\gamma\delta} \cdot \tilde{\varphi} \cdot {_{i+1}A_x} \cdot {_iB_\beta} \cdot {_iC_\gamma} \cdot {_iD_\delta}$$

$$_{i+1}A'_\alpha \text{ leading to } _{i+1}\tilde{y}_{\alpha\beta\gamma\delta} = n^*_{\alpha\beta\gamma\delta} \cdot \tilde{x} \cdot {_{i+1}A_x} \cdot {_iB_\gamma} \cdot {_iC_\gamma} \cdot {_iD_\delta}$$

With these improved values of $_{i+1}\tilde{n}_{\alpha\beta\gamma\delta}$ and $_{i+1}y_{\alpha\beta\gamma\delta}$ the same adjustment is made in β, then in γ, then in δ. After a *cycle of four steps* $= i + 4$, the approximation is much improved.

Starting from "all factors $= 1$", calculations made by Hollerith cards will generally be complete in two or three cycles, as further adjustments are irrelevant.

This definite factor system is utilized for further structure analysis:

(1) *The fitting is tested by the method of least squares applied to all minor groups in $(\alpha \beta \gamma \delta)$ and also to some major groups.*

Previously, I have proved *that the relative error in numbers n^** is expressed by $(1 \pm 1/\sqrt{n})$ and *relative error in amounts y^** by $(1 \pm v/\sqrt{n})$, where for branches with small uniform risk sums $v \gtrsim 3$ and for branches where a large proportion of claim amount derive from very large claims $v \sim 6$. Utilizing the corresponding approximate variances *the χ^2 measure can be calculated.*

(2) *By the system of calculation the following equations are satisfied:*

$$n^*_\alpha = \tilde{n}_\alpha \qquad n^*_\beta = \tilde{n}_\beta \qquad n^*_\gamma = \tilde{n}_\gamma \qquad n^*_\delta = \tilde{n}_\delta$$

$$y^*_\alpha = \tilde{y}_\alpha \qquad y^*_\beta = \tilde{y}_\beta \qquad y^*_\gamma = \tilde{y}_\gamma \qquad y^*_\delta = \tilde{y}_\delta$$

The goodness of fitting is further tested by making the same *comparison on the two-parameter margins* $\alpha\beta \; \alpha\gamma \; \alpha\delta \; \beta\gamma \; \beta\delta \; \gamma\delta$.

This test is *most important*, because it gives a reliable *answer to the question if risk structure is multiplicative* and *if a factor tariff is equitable.* (If not, some few parts of the risk system are often approximately multiplicative. Analysis is often improved if ordinary claims and excess (=large) claims are analysed separately).

(3) The factor system calculated *for several years* (δ) should furnish *a very good representation of the a priori risk.* Starting from this system we have *an opportunity to reintroduce stochastic tests*, which were *previously meaningless owing to the non-stationary character of the risk system.*

ELEMENTARY FACTOR ANALYSIS

Sometimes the tariff system is recently revised by statistics. In that case risk factors from the tariff hypothesis might be introduced as "risk a priory" or better *expected claim frequency and risk premium.*

Comparing these *"expected values"* for all risk groups $(\alpha\beta\gamma\delta)$, with *claims occurred* we have an additive system, and *although risk for minor groups is uncertain, the risk for different major groups within* $(\alpha \; \beta \; \gamma \; \delta)$ will lead to *acceptable tests.*

This method, however, is somewhat *double-edged as non-occurred large claims might have a large influence on the comparison.*