Actuarial Statistical Models - An Elementary Review

By S. Chidambaram

Abstract

Elementary statistical methods are touched upon to describe data emanating from insurers. Without going deeply into a specific model, the paper describes the methodology to choose and fit a model. The merits of different methods are also explained.

Key words
Models, Lagrange's formula, Maximum Likelihood Method, Simulation, Bayesian estimates

1. Introduction:

The principal problems with regard to pricing which confront an actuary especially in Non-Life sector are three fold. First, the actuary has to evaluate the risk, secondly, he has to adjust the evaluated risk to conform to contractual terms and thirdly he has then to arrive at the premium at which the product can be allowed to be in the market. The evaluation of the risk which is the basis of the whole exercise has to be done addressing two other aspects: the evaluation process should pay enough regard to the known claim experience and also amenable to further easier analysis and adjustments dictated by developing changes in the risk profile. In this paper an attempt is made to introduce the available methods and then discuss briefly the merit or otherwise of these methods in the context of a few specific insurances. Some illustrations are also made using hypothetical data. Since the data used are hypothetical needless to add that it cannot become directly applicable to any type of specific insurance problem. At best it suggests a way to address the problem for which the practitioner has to use relevant and reliable data drawn from actual experience.

2. Evaluation:

We are interested in knowing: How frequently the event associated with a given risk will happen during an insured period? Again, given the event has happened what would be the size of the loss associated with it? For example, suppose we are examining the risk associated with automobile accidents leading to third party liability. First, we are interested in knowing how many accidents can be expected in the insured period. Again once there is an accident what would be the likely size of the third party liability? Thus the risk evaluation process has to determine the distribution of the probability associated with the number of accidents and then the probability of a loss not exceeding a given size will also have to be settled. The number of accidents is per exposure is generally referred to as the frequency of the risk and the size of the loss arising from any accident is called the severity of the risk. Clearly the value of the risk depends both on its frequency and its mean severity.

How do we measure the frequency? Practice differs from risk to risk and company to company. In an automobile accident risk the frequency can be the ratio of the number of accidents in a year to the number of insured cars exposed to the risk. Here the accidents counted will include more than one accident resulting from a single car during the insured period. In passenger Airline accidents, the frequency may be number of accidents per passenger mileage where the exposure unit is the number of passenger mileage. Thus the manner in which is frequency is determined and evaluated can be different. The pure premium for the risk is thus the product of the frequency and the mean severity. For example,

Frequency = (Number of automobile accidents)/(Total number of cars insured)
Mean Severity = (Total amount of loss incurred from accidents)/(Total number of accidents) Clearly,
Pure premium = Frequency x Mean Severity

Whatever is our concern, evaluation of the frequency or size of the loss, the starting point is the data available to the actuary. Suppose the data for a given year for the automobile insurer are as follows:

Total Number of Cars insured = 12489
Total Number of accidents that occurred = 78

Let the loss incurred during a year from each accident be as given in the table below:

Size of the Loss	Number of accidents	Total Loss
Less than Rs.500	28	11480
Rs500 or less than Rs.2500	18	14850
Rs.2500 or less than Rs.7500	10	32500
Rs.7500 or less than Rs.10000	8	64800
Rs.10000 or less that Rs.15000	8	101600
Rs.15000 or less than Rs.20000	5	86250
Rs.20000 or Rs.25000	1	24300
Total	78	335780

The above data tells us about the number of accidents, but there may be some cars that might have produced more than one claim. Similarly there may be claims which may be less than 100 for which the policyholder may not be induced to make a claim. Thus, the data is incomplete in the sense that all losses have not been reported. Again the claims need not be accurate and the settlements might have rounded off the claims to extent. So, the amount of loss is not completely reflected. So we have only data which is per se impure by not fully accounting all accidents and only gives the claims made which need not be the actual loss.

From the above data, the frequency = 78/12489 =0.006245
Mean Severity = 335780/78 =4304.87
Pure Premium = frequency x severity = 26.88 per car

The pure premium we arrived at is simply the result of what data got reported and what amounts got settled as claims and so the true nature of the accident rate and the severity of loss are quite different. Clearly, we have to do some further operations on the available data before we can conclude that the pure premium resulting from such modified data can be reasonably taken as reflecting the future prospects. Since the basis of any rate making operations is the data, it is essential that the purity and completeness of the data are ensured. Where we know that some part of the data can be seen as spurious or where some gaps in information exists, some way must be found to purify or clean the data and fill in the missing data before the whole set can become serviceable for our purpose.

3. Limitations:

The insurance cover granted is regulated by several kinds of limitations. There is a natural limit dictated by law that ensures that the insured's claim is limited by the principle of indemnity.

Insurance cannot be a means to make gains by those who seek cover. However in every contract there is an upper limit up to which compensation is paid and beyond that the insured has to absorb the loss himself. Since processing of claims involves expenses which is more or less constant with respect to the about of the claim, it is in the interests of the insurer to put a minimum limit below which no claims are admitted by contract. These lower limits are referred to as deductibles. There is what is called franchise deductibles under which the insured has to bear the loss himself if the loss is below this limit. If however the loss exceeds the limit the entire loss becomes the responsibility of the insurer. Under vanishing deductibles if loss exceeds the limit only a percentage of the deductible is met by the insurer and this percentage decreases with increase in size of the loss and beyond a certain amount the deductible is totally ignored.

The effect of the limits is to somewhat distort the amount claimed. When the loss is near about the deductible, there is a tendency to slightly overstate the loss so that the claim exceeds the franchise limit. Similarly the maximum also distorts the information on loss distribution. When the loss exceeds the maximum, there is a tendency not to be so concerned about the determination of the actual size of the loss.

Sometimes limitation operates with the introduction of a waiting period for benefit to commence as in sickness insurance. There can be tendency to prolonging the illness beyond the waiting period so that a claim can be preferred.

So limitations play an important role in the quality of data accumulating in an office. While studying data for its accuracy, it is also necessary to understand the effect of any limitations so that the data are reorganized to eliminate or at best minimize such distortions.

All limitations in substance eliminate a certain proportion of the overall loss. An idea about this Loss Elimination Ratio (LER) is useful for a person studying the behaviour of a Loss. LER is simply the ratio of the total mount of losses eliminated to the total losses actually incurred. Limitations eliminate the severity of losses and so LER is also equivalently the ratio of severity of losses eliminated to total severity.

4. Data and Models – numerical/graphical methods:

Apart from the distortions introduced by the operation of limitations, data can become distorted due to clustering. Loss adjusters tend to settle losses at or below their authorized limit and there is also a tendency to prefer round figures. The raw data before an actuary have somehow to be modified so that these kinds of distortions are removed. One way is to identify the cluster points and then make groupings of the data in such a way that these cluster points lie in the middle of a group.

Models describe the pattern in which the losses from a risk materialize. A loss is looked upon as a random variable and the occurrence of loss values is assumed to follow a probability distribution. We can classify models as either empirical or mathematical. Under empirical models we attempt to construct models whose probability distribution cannot be described by any mathematical format. The value of the random variable and its occurrence probability are described in a tabular form estimated from the available data. On the contrary there are several mathematical models associated with random variables which neatly conform to actual experience. In insurance practice both empirical and mathematical are important.

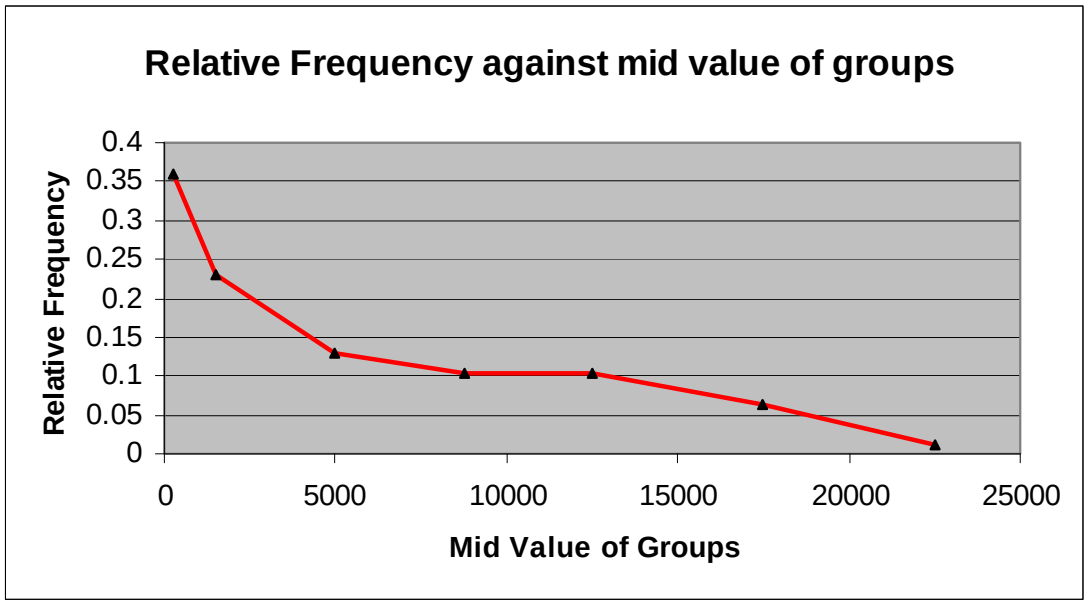
A risk if it can be modeled satisfactorily according to some known mathematical probability distribution, it can be handled with greater ease, for it is amenable to further mathematical

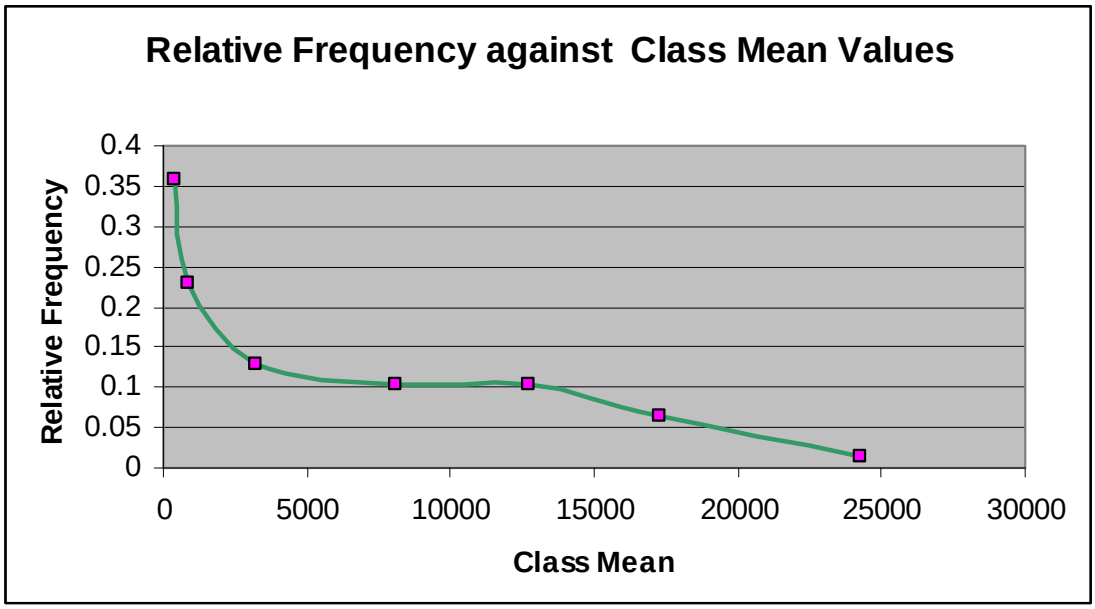
treatment and statistical principles. The elements of subjectivism in the decision making can be kept down to a minimum. But though many insurance losses are known to conform to known statistical models, their alignment is only at best close and not exact. Hence, even with carefully chosen mathematical models the elements of subjective judgment remains to an extent. The great advantage of a mathematical model is that once a good fit is achieved, the data can be practically disregarded and further problems associated with evaluation of risk is a matter of statistical mathematics. As and when new data get accumulated, the model in use can be further tested against the fresh data to see any further adjustment has become necessary. Such adjustments can be done by fine tuning the parameters of the model being used.

The empirical model certainly firmly rooted on the data in hand. When newer data get accumulated and the data volume increases it can be argued that the empirical model tends to become more and more reliable. But if the risk profile is also gradually changing and these changes are best reflected only in the more recent data, what is the relevance to include old data, in any case data belonging to the distant past? Should we discard old data in preference to newer data or should we continue to use both in the valuation of the risk? On the contrary if an acceptable mathematical model is already in place will it not be more simple and advantageous to continue using the model and give effect to the impact of newer data by adjusting the parameters of the distribution?

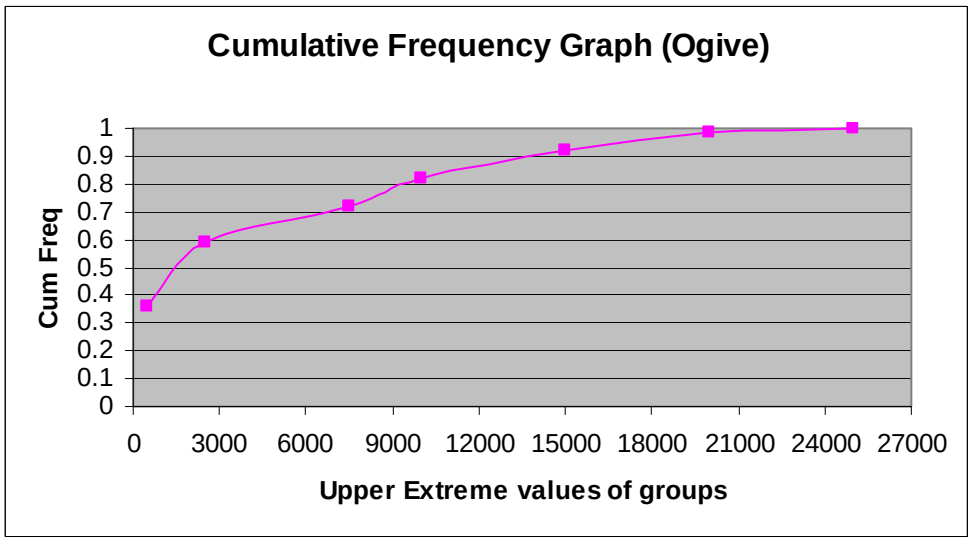
The answer to these questions is not simple. The element of subjectivism will always be present and there is always a role for informed judgment. Visualizing the future from what has happened in the past will always contain an error term. So, the success of the evaluation process lies not only in choosing the right model, empirical or mathematical, but also in the understanding of the error term inevitably involved in such exercises. We should allow for this error term in our evaluation to a sufficient degree.

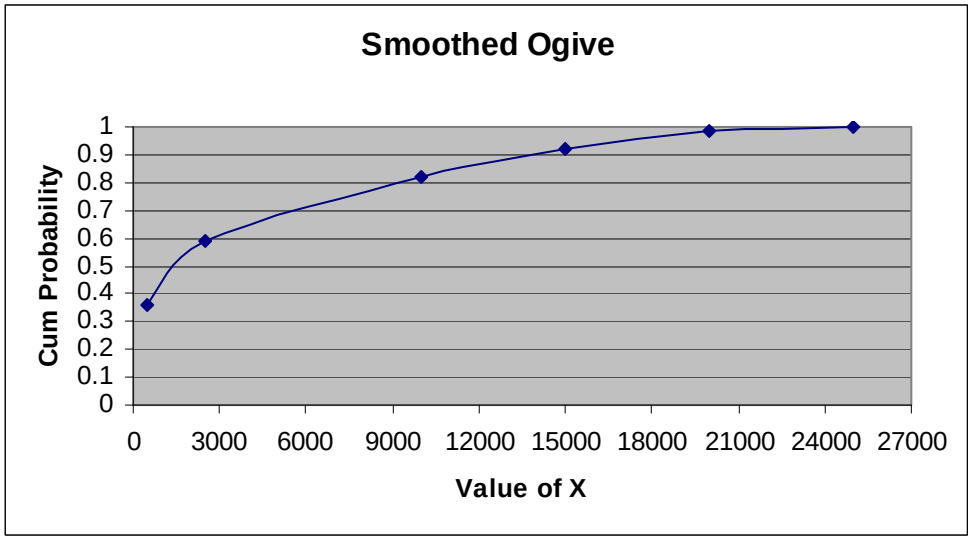
We shall now attempt to process the data given in the above table to fit an empirical distribution. As a first step we shall examine the plots of relative frequencies. The representative value for each class can be either the mid value of the class or the mean value for the class if more information is available. Since the aggregate loss details for each class are available we can use class mean also. Plots over mid-x and mean-x have been constructed using the available data.





The plot against class means appear better than the plot against the mid values of the class intervals. But the relative frequency in both needs some adjustments to conform it to a probability distribution. To do this we shall construct the ogive and look through that graph.





In the unsmoothed Ogive the curve takes concave form between x=6000 and x=12000 which means that the curve is not an increasing function and so it needs adjustment and in the smoothed form this anomaly has been rectified by inspection and regrouping of the classes. Using the smoothed ogive H(x) we can estimate the cumulative distribution function values at various points of x by reading out such values from the smoothed ogive. It is also possible to establish an approximate mathematical form to F(x) by applying finite Lagrange's Formula. From the observed data as arrived at above we choose four values to fit a polynomial for the Cumulative Distributive Function F(x). The four values chosen are highlighted in the table below below:

Group Upper Extreme (x)	(y)	Cumulative Frequency F(x) or F(y)
500	**0.5**	**0.35897436**
2500	2.5	0.58974359
7500	**7.5**	**0.71794872**
10000	10	0.82051282
15000	**15**	**0.92307692**
20000	20	0.98717949
25000	**25**	**0.99999999**

In order to make the numerical work simpler the variable X has been scaled down by a factor of 1000 and these scaled down variable Y has also been shown in the above table. The form of F(x) can now be determined using Lagrange's formula which has its origin in the theory of divided differences.

Now, we have

$$\frac{F(x)}{(y-0.5)*(y-7.5)*(y-15)*(y-25)} =$$

$$\frac{F(0.5)}{(0.5-7.5)*(0.5-15)*(0.5-25)} * (1/(y-0.5) +$$

$$\frac{F(7.5)}{(7.5-0.5)*(7.5-15)*(7.5-25)} * (1/(y-7.5) +$$

$$\frac{F(15)}{(15-0.5)*(15-7.5)*(15-25)} * (1/(y-0.5) +$$

$$\frac{F(25)}{(25-0.5)*(25-7.5)*(25-15)} * (1/(y-25)$$

By simplifying the above equation we arrive at the relationship

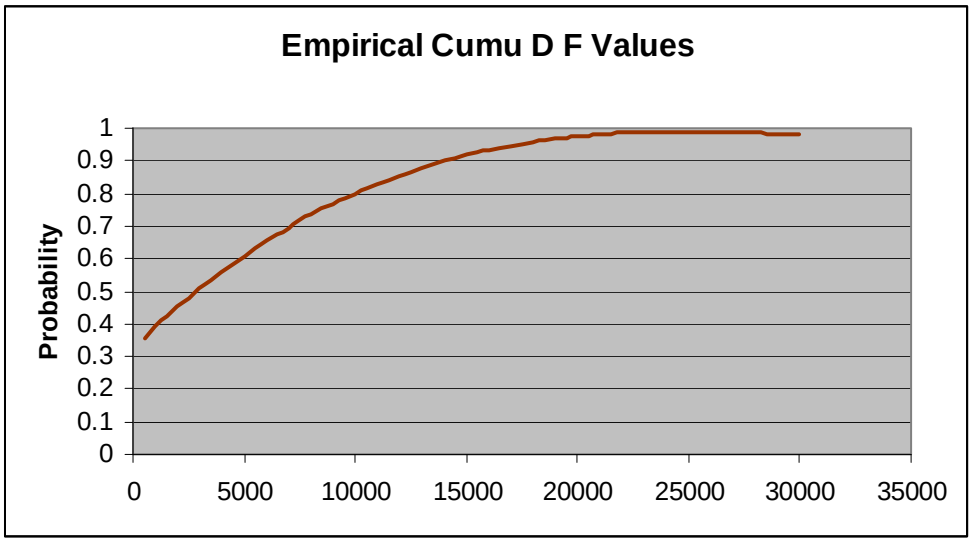
$$\mathbf{F(y) = 0.0000215 * y^3 - 0.0021453 * y^2 + 0.0667576 * y + 0.3259339}$$

We can now use the above relation to determine values of F(x) for any value of X remembering that x=1000 * y. The calculated values are tabulated below:

Empirical Values of F(x)

(x)	F(x)	(x)	F(x)	(x)	F(x)	(x)	F(x)
500	0.358779	8000	0.733710	15500	0.925385	23000	0.988266
1000	0.390568	8500	0.751587	16000	0.932982	23500	0.989213
1500	0.421316	9000	0.768666	16500	0.940022	24000	0.989845
2000	0.451040	9500	0.784963	17000	0.946522	24500	0.99018
2500	0.479756	10000	0.800493	17500	0.952498	25000	0.990232
3000	0.507480	10500	0.815274	18000	0.957966	25500	0.990019
3500	0.534228	11000	0.829321	18500	0.962943	26000	0.989555
4000	0.560016	11500	0.842650	19000	0.967444	26500	0.988859
4500	0.584861	12000	0.855278	19500	0.971485	27000	0.987945
5000	0.608778	12500	0.867220	20000	0.975083	27500	0.986829
5500	0.631784	13000	0.878493	20500	0.978254	28000	0.985529
6000	0.653895	13500	0.889113	21000	0.981014	28500	0.98406
6500	0.675127	14000	0.899096	21500	0.983379	29000	0.982437
7000	0.695496	14500	0.908459	22000	0.985365	29500	0.980678
7500	0.715018	15000	0.917216	22500	0.986989	30000	0.978799

It will also be interesting to view the graph of F(x) using the empirically determined values. The graph is shown below:



After arriving at the distribution as aforesaid, it is necessary to check the distribution to ascertain whether the basic laws of probability are violated by the distribution. The two principles to be checked are that the d.f is an increasing function and that at the maximum value the variable can take the d.f. must not exceed 1. If any anomaly is noticed, the entire exercise will have to be redone with a different grouping of the data which the experimenter believes would remove the anomaly. To trace how closely the resulting distribution follows the data it is possible to apply a Chi-Square test of goodness of fit between the actual data and the resulting corresponding figures from the numerical exercise done. The success of the above suggested method is not always assured. If the data are too complicated a good fit may not result.

5. Data and Models – Simulation

The numerical method suggested in the previous section may not be convenient in many situations. The work load involved may be too heavy or the resulting distribution may not conform to the laws of probability distribution however much you try to adjust the resulting distribution. In such situations a simulation method is a possible way out.

Here we attempt to fit a known distribution to the random variable. From prior knowledge the broad shape of the distribution might be known and that may be the reason for the choice of simulation. A simulation can be attempted using a uniform distribution. The rationale for the simulation is as follows.

Let Y have a uniform distribution, where the support is $0 <= y < 1$. Let X be the variable which we want to simulate. Let Y= F(X). X the random variable is thus defined as the inverse function of Y.

That is
$X = F^{-1}(Y)$. So the distribution function of X will be:

$$\Pr[X <= x] = \Pr[Y <= F(x)] = \Pr[F(X) <= F(x)]$$

Since $Y = F(X)$ we have

$$\Pr[X=x] = \int_0^{F(x)} (1)dy = F(x).$$

Thus X values can be derived from any randomly chosen value of y from the uniform distribution using the inverse relationship of X with Y as defined above. We can thus generate any number of random values for X from random numbers taken from the uniform distribution. We shall illustrate this with an example. Suppose that we want to develop a Pareto distribution for X from a uniform distribution.

We know that the d f of a Pareto distribution is given by

$$F(x) = 1 - \lambda^{\alpha} (\lambda + x)^{-\alpha} \qquad 0 <= x < \infty$$

We set $Y = 1 - \lambda^{\alpha} (\lambda + x)^{-\alpha}$
so that

$$X = \lambda [1/(1-Y)^{1/\alpha} - 1]$$

Now using a computer we pick out a random variable and then determine the corresponding value of X from the above relationship. But we must have knowledge of the two parameters λ and α .
These we derive from the available data as the first approximation. Let the available data be as follows:

Class Interval	Mid-x (x)	Frequency (f)	f*x	f*x^2
0-500	250	28	7000	1750000
500-2500	1500	18	27000	40500000
2500-7500	5000	10	50000	250000000
7500-10000	8750	8	70000	612500000
10000-5000	12500	8	100000	1250000000
15000 - 0000	17500	5	87500	1531250000
20000-5000	22500	1	22500	506250000
Total		78	364000	4192250000
Mean E(X)	4666.67			
E(X^2)	53746794.87			

For Pareto distribution we set $\alpha = 2.\ (m2 - m1^2)/(m2 - 2m1^2)$

where m1= 4666.67

and m2= 53746794.84 and $\lambda = m1*m2/(m2 - 2m1^2)$

Hence $\alpha = 6.2738233$

$\lambda = 24611.175$

Using these initial values of the two parameters we simulate from a computer program a set of values, say 500. Now we make a frequency distribution out of these 500 values and calculate again the new parameter values. We repeat the simulation using these new parameter values to produce another set of 500 random numbers. The process is again and again repeated to refine the parameter values further. In this way we arrive at a set of F(x) for the Pareto distribution by applying the final refined parameter values.

In the above example, we were able to derive d f of X in a closed form and we derived several sets of random values for X only to refine the parameter values. But it is not generally possible to arrive at the functional form of the d f of X that simply because of the complexity in solving for X from the relation $X= F^{-1}(Y)$. The values corresponding to the Random Y-values from uniform distribution are determined by some numerical methods. However, with a suitable computer program this laborious work can be lightened.

Before concluding this part we shall indicate how greater complexity arises by assuming that X has a Gamma distribution in stead of a Pareto as earlier assumed. With Gamma we set

$$Y = (1/\Gamma(\alpha))* \int_0^x (\beta/\Gamma\alpha)* \exp(-\beta x)* (\beta x)^{\alpha-1}\, dx$$

Hence the random values derived from the uniform distribution will follow a Gamma distribution with

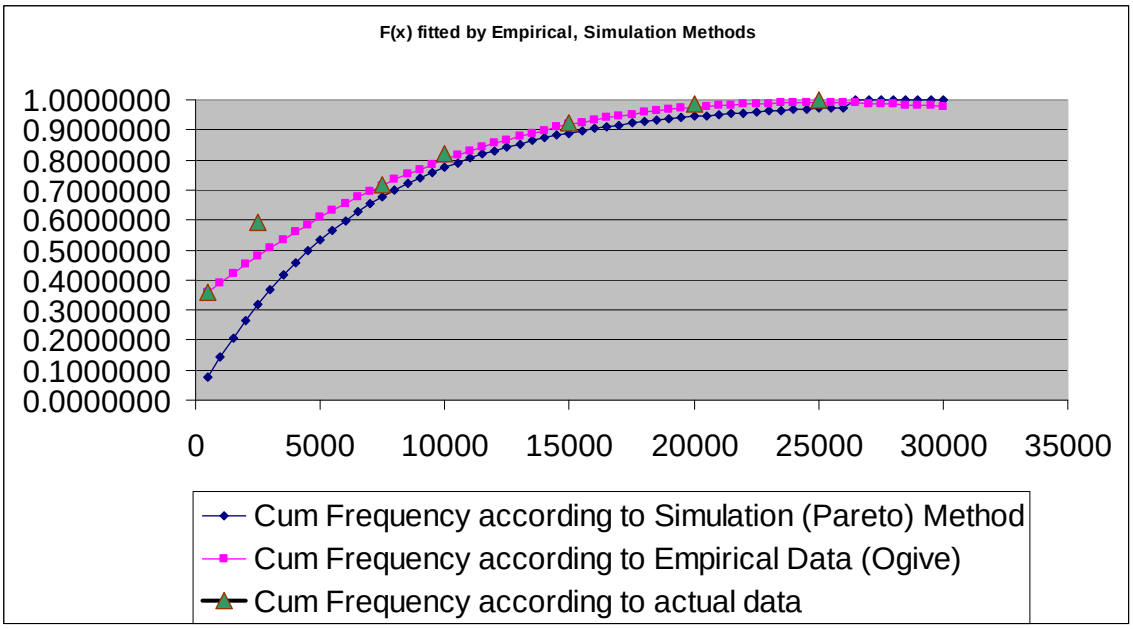
Mean $= \alpha/\beta$ and

Variance $= \alpha/\beta^2$

The starting parameter values are derived from the data available by equating the data mean and data variance to the above relations.

But the calculation of X values from the obtained random values of Y can now be seen to be quite daunting, not to mention the evaluation of integral by numerical methods to get F(x) values. Considerable programming skill and computer time are required to do this.

Random variables of the type Pareto were generated using a uniform distribution from a computer and these random Pareto values were grouped and then their cumulative frequencies were determined and tabulated. Every time 500 sets of random values were derived and from that the parameter values of the Pareto were recalculated. The trial was repeated for another set of 500 random values using the new parameter values derived. Two hundred five such trials were made and the mean parameter values resulting from these trials were then used to produce the F(x) values. These final F(x) values derived are tabulated below:



Cumulative Distribution Function Values for the Pareto with Alpha =20.33 and Delta =131024

Compared with the fitted empirical model and actual data –(Graph shown above)

Upper Bound of Class	Cum Frequency according to Simulation (Pareto) Method	Cum Frequency according to Empirical Data (Ogive)	Cum Frequency according to actual data	Upper Bound of Class	Cum Frequency according to Simulation (Pareto) Method	Cum Frequency according to Empirical Data (Ogive)	Cum Frequency according to actual data
500	0.074511	0.358779	0.358974	15500	0.897005	0.925385	
1000	0.143219	0.390568		16000	0.903896	0.932982	
1500	0.206595	0.421316		16500	0.910306	0.940022	
2000	0.265070	0.451040		17000	0.916268	0.946522	
2500	0.319040	0.479756	0.589744	17500	0.921816	0.952498	
3000	0.368867	0.507480		18000	0.926980	0.957966	
3500	0.414882	0.534228		18500	0.931787	0.962943	
4000	0.457390	0.560016		19000	0.936263	0.967444	
4500	0.496669	0.584861		19500	0.940431	0.971485	
5000	0.532976	0.608778		20000	0.944315	0.975083	0.987179
5500	0.566545	0.631784		20500	0.947934	0.978254	
6000	0.597592	0.653895		21000	0.951307	0.981014	
6500	0.626313	0.675127		21500	0.954451	0.983379	
7000	0.652891	0.695496		22000	0.957383	0.985365	
7500	0.677493	0.715018	0.717949	22500	0.960118	0.986989	
8000	0.700272	0.733710		23000	0.962669	0.988266	
8500	0.721368	0.751587		23500	0.965049	0.989213	
9000	0.740913	0.768666		24000	0.967271	0.989845	
9500	0.759023	0.784963		24500	0.969345	0.990180	
10000	0.775810	0.800493	0.820513	25000	0.971282	0.990232	1.000000
10500	0.791375	0.815274		25500	0.973090	0.990019	
11000	0.805809	0.829321		26000	0.974780	0.989555	

11500	0.819199	0.842650		26500	1.000000	0.988859
12000	0.831624	0.855278		27000	1.000000	0.987945
12500	0.843156	0.867220		27500	1.000000	0.986829
13000	0.853862	0.878493		28000	1.000000	0.985529
13500	0.863804	0.889113		28500	1.000000	0.984060
14000	0.873039	0.899096		29000	1.000000	0.982437
14500	0.881619	0.908459		29500	1.000000	0.980678
15000	0.889593	0.917216	0.923077	30000	1.000000	0.978799

6. Applying chosen probability models to a given set of data:

We have considered two approaches to finding the distribution function of random variable, the empirical and the simulation methods. In the empirical method an ogive was determined in a mathematical form using the available data. In other words a polynomial was fitted to F(x) from the observed values. Once the polynomial is thus determined we could generate F(x) values from any number ox x-values. Here we could say that we arrived at our distribution function without assuming much about the shape of the model. The data was allowed to determined the functional structure of F(x). Under the simulation technique we assumed that the data can be expected to follow a known distribution. In our example we assumed it to be a Pareto. Then using the technique of generating

random numbers from a uniform distribution we could derive several random numbers applicable to the given data. These random values (which will conform to the assumed Pareto) were used to produce a fresh set of data from which the parameters of the Pareto were re-estimated. Again the process was repeated several times using every time the fresh set of parameters just previously determine. We thus get a set of parameter values for the Pareto we assumed for the data. Using the mean value of this set of parameter values, we generated the simulated F(x) values. Here again the method was empirical but there was some assumption about the likely type of distribution.

In this section we shall consider an approach which totally assumes that the given data originated from a known statistical distribution. As we know any statistical distribution will have a few parameters that determine the shape and structure of the distribution function. Since we have assumed here the statistical distribution, the only problem that remains is to determine its appropriate parameter values. The available data are then used to estimate these parameter values. The technique clearly assumes much more than the two methods described above. Even though in the simulation method we assumed a distribution, the parameter values were determined by a simulated random process. When we later discuss the relative merits of these methods the difference of approach will become clearer. Under the model based approach there is not only the assumption that the distribution is pre known but the parameter values of that distribution can be derived from the data itself by some method. The assumptions as to the distribution are much more than in under simulation technique. In other words we strongly believe that the distribution function applicable to the given data and its population follows the assumed model and therefore the model is sacrosanct.

The estimation of parameter values can be done by different statistical approaches. One way is by the maximum likelihood and another is the comparison of the moments of the random variable. Where the assumed model provides a theoretical means to express the parameter values in terms of the moments of its distribution and if we have some sample data also available, then the sample moments can be used as a first estimate of the population moments and in this way the parameter values estimated. There may more adjustments needed dictated by any further knowledge available

about the risk and attempts should be made to carry them into our first estimate to produce a better estimate.

As for maximum likelihood estimate of the parameter values, we start by defining a Likelihood function involving the parameters to be estimated. Let f(x) be the theoretical p.d.f. associated with the assumed model. Then if we have n sample values of X, we make use of these n sample values to construct our likelihood functions as:

$$L(\alpha,\lambda) = \prod_{i=1}^{i=n} [\, f(x_i) \,]$$

In the case of Pareto distribution (assumed model) since we have $f(x) = \alpha\lambda^{\alpha} (1 + x)^{-(\lambda+1)}$

$$Ln\ (\alpha,\lambda) \quad = \quad n[Ln(\alpha) + \alpha\ Ln(\lambda)] - (\lambda + 1) \sum_{i=1}^{i=n} Ln(1 + x_i)$$

By differentiating the above function with respect to the parameter λ (treated here as a variable and the other parameter as constant) we obtain a relationship involving the two parameters. Similarly by differentiating with respect to the other parameter α (treating it as a variable and treating λ as a constant) we get another relationship involving the two parameters. If we set the two derivatives to zero and assume that the second derivatives are negatives, then we may be able to determine numerical values for the two parameters. After determining the numerical values we may test the second derivatives to ascertain if they do in fact yield negative values with these numerical values. If not, the estimates we arrived at may not be suitable and some other methods have to be looked for.

In certain models with a single parameter, if we have a strong belief that the parameter of the distribution itself can be treated as a random variable having a standard mathematical form then we may further assume that the likelihood function is the conditional joint probability distribution of the n random variables for a given parameter value. In such circumstances, the joint pdf of these n random variables and the parameter can be taken as the product of the likelihood function and the pdf of assumed for the parameter. The probability distribution of the parameter when the n random variables are known to assume the observed values will be determined by summing this joint pdf of the n random variables and the parameter over the range in which the parameter can exist.

Where the number of parameters is more than one, the above Bayesian method will be successful only if some further assumption can be made about their inter-relationship. For instance if the parameters are λ and α , then if it can be assumed that $\lambda = k\alpha$, where k is a constant, the distribution can be converted into a single parameter distribution and the above method adopted.

7. Comparison of model based distributions

We had touched upon how we would go about fitting a distribution on the basis of an assumption of a mathematical model distribution. In all the three set of cumulative frequencies, the distribution is pre-assumed and fitted by choosing an appropriate value for the parameters of the resulting distribution. The simplest is finding the parameters by the method of moments. Then we touched upon the maximum likelihood estimates of the parameters and then also on how Bayesian estimation of parameter is possible. Which is the best? Of course it is that which lies nearer the true value. One way of comparison is to estimate the error of our estimate under each method. The variance even if it can be approximately estimated provides some clue. Which one will converge more rapidly to the true value can then be gauged from the variances.

It has been found that with large samples the method of maximum likelihood is superior to the method of moments. When it comes to Bayesian estimates, the error will be even lower provided we could somehow hit upon the prior distribution which describes the parameter in the right region.

Whatever the model is chosen, it should be tested for goodness of fit with the observed data.

8. Conclusions

The process of statistical modelling begins with available data, or only collateral data or no data. The value of the risk being modelled need adjustments because the loss actually contracted to be met will almost certainly be different from the actual progression of the value of the risk. Usually the data available for modelling is data on loss and not on the actual incurred value of the risk. A successful model will be one that can capture the actual value of the risk so that when the contracted loss has to be varied the evaluation of the contracted risk would be appropriate.

When we have sufficient data, the exercise would be more reliable. The type of model would be dictated by the available information on the pattern of the distribution function of the risk and its parameter values. When only collateral data are available, then the model will have to be more prudently cast so that the risk is not underestimated. The area of interest in the body of the risk is important. For example if there is to be a maximum limit for the loss the distribution should be adjusted so that the probability of occurrence of the loss within the band of interest is not understated. This would be the case with most exercise with rate making. On the contrary if the model is to serve for reserving, or fixing reinsurance limits, then a different distribution of the probability resulting may be necessary with a thicker tail. Here the relative emphasis being on the tail, we reconstruct the model so that a larger part of the probability lies in the tail compared to the originally assumed distribution.

When no data is available the exercise becomes even more difficult. Pilots are required to acquire experience and the pilots have to be planned in such a way that loss as well as the size of the pilot is suitably limited so that while information on the true value of the risk is gathered, the office is not too much exposed. Based on such pilot, then the modelling process can commence in the ways indicated above.

The evaluation resulting from a modelling exercise has to be further adjusted for expenses of office and security loading and data insufficiency factors. Security loading is an addition dictated by the model so that the probability of running into a situation when the accumulated premiums plus the loadings do produce an excessive aggregate loss, the probability of ruin, is an acceptable low for the office.

The points expressed in this paper are merely indicative of the possible means available to actuaries and the illustrations shown above do not apply to any specific situation of an office. In short, it might be looked upon as a source of broad methods, but the success of any modelling exercise will depend upon the volume of data available, the right choice of the mathematical model and sufficient loadings for security etc. The emerging final office rates may further be dictated by competitive pressure. However a prudent actuary will take enough steps not to be swerved by such pressure and should work the final product with clear and easily recognizable limitations and terms so that the resulting modified rates still conform to his original findings.

References:

1. Loss Distributions By Robert V Hogg & Stuart A Klugman

2. Introductory Statistics with applications in General Insurance By I B Hossack, JH Pollard, B Zehnwirth

3. General Insurance By Insurance Institute of India Practice of General Insurance By Insurance Institute of India

4. Risk Theory By Newton, L Bowers, Gerber, Hickman, Jones & Nesbitt

About the Author:

S. Chidambaram

- Running an independent actuarial consultancy service tendering advice on matters relating to Life insurance, Pension and Employee Benefit Risks
- Valuation of Gratuity Funds, estimation of Leave Encashment Benefit costs, designing Pension Schemes and other Employee Benefits Schemes by applying statistical theory and methods
- Advising Govt of Kerala on actuarial matters including valuation and product design of their official branch life portfolio
- Statistical Analysis for various research data and reporting on results – Eg. Regional Cancer Centre, Tropical Botanical Gardens , Sixth Pay Commission etc
- Appointed Actuary for Agriculture Insurance Co. of India, New Delhi