Reinventing Risk Classification—A Set Theory Approach

Romel G. Salam, FCAS, MAAA

Reinventing Risk Classification – A Set Theory Approach

Romel G. Salam

Abstract

Risk Classification represents one of the most important and controversial topics of actuarial science. It is covered broadly throughout the Casualty Actuarial Society's exam syllabus. The importance and persistence of this topic is also reflected in the long array of papers that permeate the casualty actuarial literature. Most of the recent work on Risk Classification has focused on automobile insurance coverage, which is principally responsible for bringing the issue into the public debate. However, Risk Classification impacts on all types of insurance coverage and has ramifications beyond the world of insurance.

Risk Classification starts necessarily as a subjective process. The characteristics along which risks are delineated are intuitive at best. Traditional treatments of Risk Classification in the actuarial literature, in our view, do not provide the tools to move beyond intuition. In this paper, we will review the common definitions of Risk Classification by quickly glancing through two reference materials on the subject: the American Academy of Actuaries Risk Classification Statement of Principles [2] and Robert Finger's chapter on Risk Classification in the Foundations of Casualty Actuarial Science textbook [6]. Then, by building on the existing definitions, we will look to establish a more rigorous and consistent treatment of the subject. At the core of our treatment will be a non-traditional definition of the notion of *class*. We will borrow terminology from Set Theory[1] to help us in this endeavor. We will not only define more rigorously such concepts as *homogeneity* and *separation* but we will also integrate them into the very definition of Risk Classification. A method of Risk Classification will emerge as a natural byproduct of our definitions. This method, which may be described as what Venter [10, p. 345] terms a "credibility only" method, will provide an alternative to using arithmetic functions in Risk Classification schemes. To illustrate our newly

[1] Familiarity with elementary Set Theory, although not required, is helpful in order to understand the material presented herein. For an introduction to Set Theory, see Gilbert and Gilbert [7].

defined precepts of Risk Classification, we will construct a specific model using simulated observations. We will introduce a set of statistics that will allow us to make inferences about our model. Also, we will propose measures for assessing the relative efficiency of competing schemes and suggest procedures for validating a classification scheme. Finally, we hope that this paper will provide ideas to actuaries looking to build a Risk Classification scheme from scrap.

Introduction

Let us introduce an example where rates are being sought to provide professional liability coverage to actuaries. A *classification scheme* is proposed, which groups actuaries based on two criteria or *classification variables*: "area of practice" and "years of experience." "Area of practice" is subdivided into two (mutually exclusive) bands or *risk characteristics*: Life, Non-Life while "years of experience" is subdivided into two (mutually exclusive) *bands*: 10 or fewer, 11 or more. Four *cells* or sets of actuaries will emerge out of this arrangement: Life actuaries with 10 or fewer years of experience, Life actuaries with 11 or more years of experience, Non-Life actuaries with 10 or fewer years of experience, Non-Life actuaries with 11or more years of experience.

Why are we pooling actuaries into various cells in the first place? Couldn't we charge a single rate to all actuaries based on their combined experience? Taking this approach, we would run the risk of charging the same rates to groups that have fundamentally different loss propensities[2]. This would create subsidies across groups that carry both economic and social consequences. Conversely, are we to presume that actuaries across these cells have different loss propensities by virtue of our having separated them in this manner? Should we proceed to calculate rates for each cell based on its respective experience? Wouldn't we then run the risk of charging different rates to groups that essentially have the same loss propensity? This too might create subsidies with dire economic and social consequences. What if we had instead devised a classification scheme that grouped actuaries according to whether they were left-handed or right-

[2] Loss propensity may refer to either the probability distribution of the claim process for a cell or to the parameters and functions of parameters of the probability distributions within a cell.

handed and according to whether they sported bifocals or contact lenses (assuming all actuaries wear one or the other eye-device but not both)? Besides from lacking intuitive appeal, what separates the latter scheme from the former? Perhaps, we need to take a step back and ask ourselves what exactly is Risk Classification or its purpose. Let's look to the literature for guidance.

Current Definitions

The American Academy of Actuaries, Risk Classification Statement of Principles defines Risk Classification as "[the process of] grouping risks with similar risk characteristics for the purpose of setting prices. [2, p.1]" "Risk Classification", the Statement adds, "is intended simply to group individual risks having reasonably similar expectations of loss. [2, p.1]"

Robert Finger defines Risk Classification as the "formulation of different premiums for the same coverage based on group characteristics. [6, p.231]"

Discussion

Both the above definitions are intuitively appealing. However, in our opinion, they leave open certain key questions. For instance, the Statement's definition does not directly address the question of whether the risks across cells need to have different loss propensities. Finger, while implying in his definition that Risk Classification should recognize differences amongst cells, does not elaborate on how those differences might be recognized. The mere grouping of risks with similar characteristics, as suggested by the American Academy of Actuaries, seems like a rather incomplete goal of Risk Classification. We agree with Finger that Risk Classification must entail the emergence of differences amongst cells of risks. Otherwise, there would not be a need to classify in the first place. However, in our opinion, there should not be a presumption that any chosen risk characteristics, however intuitive, will result in cells that have different loss propensities. Before charging different rates to risks across different cells, it seems that one would need to be reasonably certain that the cells have different loss propensities. We believe that Risk

Classification should avoid two mistakes: charging the same rates to pools of risks that have fundamentally dissimilar loss propensities or charging different rates to pools of risks that have fundamentally similar loss propensities. The goal of risk classification should be to arrive at rates that closely represent the loss propensity of every risk while avoiding these the two types of mistakes.

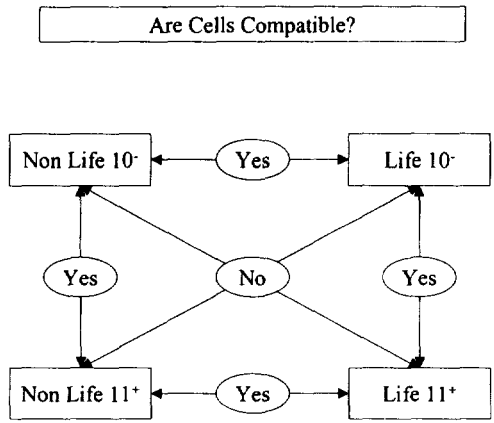
Let's return to the classification scheme for actuaries introduced earlier. We should keep in mind that it may be appropriate to devise a different classification scheme[3] for each aspect of the claim process. For instance, the rating variables along which frequency is analyzed need not coincide with those used for severity. For simplicity, let's assume we are looking at only one aspect of the claim process and that aspect alone determines the differences (if any) in the cost of coverage between cells. Let's assume a probability model is initially derived for each cell based on the respective observations in each cell. Let's finally make the assumption that the models all have the same functional form and only their parameter values may differ. Let's review the following four scenarios:

Scenario 1: In the first scenario, the parameters underlying the models for life and non-life actuaries with 10 or fewer years of experience, respectively, can't be differentiated. We will say of these cells that they are compatible[4]. Under this scenario, life and non-life actuaries with 11 or more years of experience, respectively, are also compatible. Finally, under this scenario both life and non-life actuaries with 10 or fewer years of experience, respectively, are compatible with their more experienced counterparts. This scenario is illustrated in the chart shown in figure 1 below. Has Risk Classification been successful under this scenario? Can the process even be called Risk Classification? Do any of the cells defined above constitute *classes*? More importantly, should the observations of all or any of the cells be joined for the purpose of estimating the parameters of the models?

[3] If the processes are independent as is often assumed, it makes sense to classify them separately.

[4] This narrow definition of compatibility assumes symmetry. That is, given two cells C_i and C_k, if C_i *compatible with* C_k, *this definition implies that* C_k *compatible with* C_i *and vice versa*. This will not be the case in our general definition provided later in this paper. Also, a cell is compatible with itself by definition.

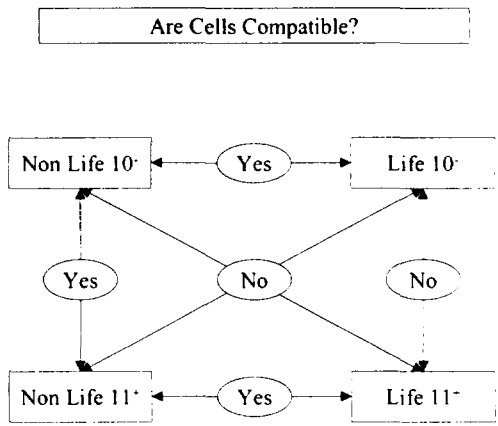
Figure 1: Compatibility[5] Chart for Scenario 1



Scenario 2: In the second scenario, it is found that the parameters underlying the models for life actuaries with 10 or fewer years of experience and those with 11 or more years of experience can be differentiated. We will say of these two cells that they are incompatible. Under the second scenario, it is found that life and non-life actuaries, respectively, who fall in the same experience group are compatible. It is also found that non-life actuaries with 10 or fewer years of experience are compatible with their more experienced counterparts. The compatibility chart is shown in figure 2 below. To what degree has Risk Classification been successful under this scenario? Do any or some of the cells defined above constitute *classes*? Should any of the cells be joined for the purpose of estimating the parameters of these models? If so, which?

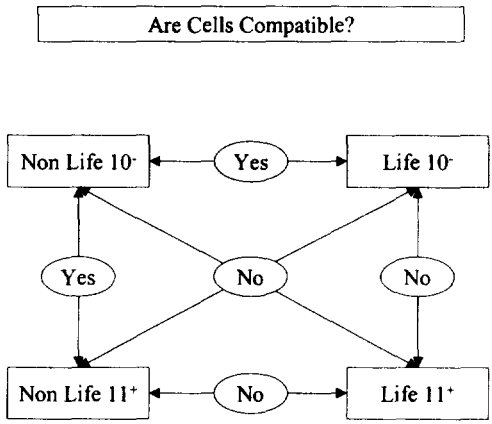
[5] Life actuaries are not compared with non-life actuaries falling in opposite experience groups, as these groups do not share any common characteristics. These comparisons would be irrelevant in the context of the given classification scheme. These pairs of cells will be defined later as non-adjacent and are incompatible by definition.

Figure 2: Compatibility Chart for Scenario 2



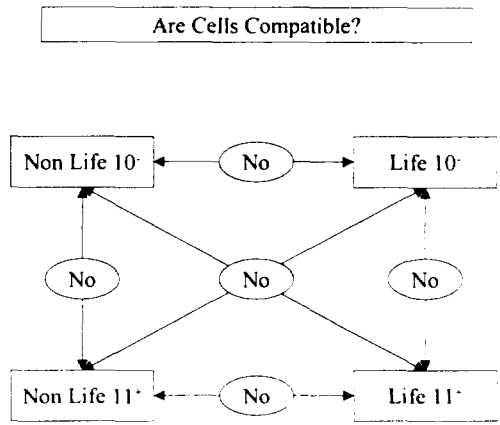
Scenario 3: In the third scenario, it is found that life actuaries with 10 or fewer years of experience and those with 11 or more years of experience are incompatible. Also, under this scenario, it is found that life and non-Life actuaries who fall in the 10 or fewer years of experience group are compatible while life and non-life actuaries who fall in the 11 or more years of experience group are incompatible. Finally, it is found that non-life actuaries with 10 or fewer years of experience are compatible with their more experienced counterparts. The compatibility chart is shown in figure 3 below. To what degree has Risk Classification been successful under this scenario? Do any or some of the cells defined above constitute *classes*? Should any of the cells be joined for the purpose of the parameters of these models? If so, which?

Figure 3: Compatibility Chart for Scenario 3



Scenario 4: Finally, in the fourth scenario, all pairs of cells are found incompatible. The compatibility chart is shown in figure 4 below. Is this the only scenario under which Risk Classification has been successful? Is this the only scenario in which the cells defined by the classification scheme constitute classes? Should any of the cells be joined for the purpose of estimating the parameters of these models? If so, which?

Figure 4: Compatibility Chart for Scenario 4



We have raised several questions in reviewing the preceding scenarios. Let's see how these questions could be answered from the perspective of the AAA's Statement of Principles and Robert Finger's chapter on the subject. Based on our understanding of the Statement of Principles, the pooling of actuaries suggested in our example would fit the AAA's definition of Risk Classification even before any of the scenarios are considered. Remember that the American Academy of Actuaries' Statement of Principles simply defines Risk Classification as "a grouping of Risk with similar risk characteristics." The Statement of Principles is silent on the issue of whether, and which cells should be joined for the purpose of estimating costs. The Statement of Principles does list credibility among three statistical considerations in designing a Classification scheme. Under this consideration, the Academy suggests that "it is desirable that each of the classes in a risk classification scheme be *large enough* to allow credible statistical predictions about that class...Accurate predictions for small, narrowly defined classes often can be made by appropriate statistical analysis of the experience for broader grouping of *correlative classes*. [2, p.10]" This implies that the parameters of a cell with a small number of observations may be estimated by joining it

with other cells, while the parameters of a cell with a large number of observations may be based on that cell alone.

Would our grouping satisfy Robert Finger's definition of Risk Classification under the first scenario? Under that scenario, the grouping would be unable to formulate statistically different premiums based on the characteristics of each cell of actuaries. What about the second scenario where only one pair of cells shows differences in the parameters of their models, or the third scenario? Would our grouping fit Finger's definition under these scenarios? Finger is also silent on the issue of whether, when, and which cells should be joined for the purpose of estimating the models' parameters. Similarly to the American Academy of Actuaries Statement of Principles, Finger mentions credibility as one of four actuarial criteria for selecting rating variables. This criterion requires that "a rating group ... be large enough to measure costs with sufficient accuracy. [6, p.237]"

The notion of credibility, as presented in Finger [6] and the American Academy of Actuaries [2] and for that matter in most actuarial papers on Risk Classification and Ratemaking, is used in what Philbrick [8, p. 214] calls "[its] familiar sense (as opposed to its technical meaning) [as] almost a synonym for confidence." "[This] terminology", Venter [11, p. 382] tells us "is misleading if it implies that the credibility weight is an inherent property of the data. " Our definition of credibility, unlike that of Finger and of the Academy, will be analogous to the technical meaning of credibility as presented in Philbrick [8, p. 214] that is credibility is " the appropriate weight to be given to a statistic of the experience in question relative to other experience."

We view the *grouping* of the actuaries into the four cells as no more than the posing of a pair of hypotheses, which roughly state:

1) Actuaries within the same cell share the same loss propensity.
2) Actuaries across different cells have different loss propensities.

We will refer to the first and second hypotheses as the *homogeneity* and *separation*[6] hypotheses, respectively. Merely setting the hypotheses does not make them true. Merely selecting classification variables and risk characteristics that seem intuitive and reasonable does not mean that the resulting cells will satisfy the hypotheses. For, intuition and reasonableness remain only subjective concepts.

Homogeneity: It may be difficult to prove directly that all risks within a cell have the same loss propensity. However, this hypothesis may be proven false if one or more risk characteristics are found such that risks within a cell can be subdivided to define new sub-cells and the risks across the newly defined sub-cells have different loss propensities. Theoretically, there are an infinite number of risk characteristics that could be used to separate risks within a cell. In reality, most potential risk characteristics are either unknown or simply unfeasible to use. Hence, one is limited to a handful of characteristics from which to choose. When a classification scheme is proposed, one may test homogeneity by introducing additional characteristics (known and feasible) to see whether the risks across the newly defined sub-cells have different loss propensities. For instance, we may introduce pension as an additional area of practice by which to pool non-life actuaries. If no such characteristics emerge, we may assume the homogeneity hypothesis to hold. Alternatively, we may simply assume that a given classification scheme provides the smallest and finest pooling of risks and no further subdivision of the cells is possible. Therefore, the homogeneity hypothesis would hold by default.

Separation: This hypothesis can be tested by successively comparing the compatibility of different pairs of cells. We assume that a test or a statistic can be devised to answer the question of compatibility between pairs of cells. For instance, given a range of values of a chosen statistic, we may conclude that two given cells are incompatible and, therefore, their parameters need to be estimated independently of each other. Conversely, for values of the chosen statistic that fall outside the range, we would conclude that two given cells are compatible. Then, the law of large numbers dictates that the observations across both cells

[6] This concept is somewhat different than the one introduced by Michael Walters who, in his 1981 Dorweiler prize-winning paper Risk Classification Standards, defines separation as "a measure of whether classes are sufficiently different in their expected losses to warrant the setting of different premium rates [12, p. 11]."

provide a better estimate of the parameters underlying the statistical models of these cells rather than just the observations in each individual cell. If all pairs of cells were incompatible, we would then accept the separation hypothesis. Then the parameters underlying each cell in the classification scheme would be estimated by relying solely on the observations from that cell. If the separation hypothesis were rejected, then one of several alternatives could be accepted. These alternative hypotheses range from finding that all pairs of cells are compatible (no need to classify at all) to finding various combinations of cells that are compatible. For example, given a cell C and a set $C_{Compatible}$ representing the union of all cells that are compatible with C excluding C itself, the estimates of the parameters of C would be derived from observations taken from C together with those taken from $C_{Compatible}$.

Credibility: When the separation hypothesis fails, the new estimates of C based on observations taken from C together with those taken from $C_{Compatible}$ can also be viewed as the credibility weighted average of the estimates based on observations from C alone with estimates based on observations taken from $C_{Compatible}$. The new estimates E_{New} of the parameters of C might then be expressed as follows:

$$E_{New} = Z \times E_{Compatible\ Cells} + (1 - Z) \times E_{Old} \qquad (1)$$

where $E_{Compatible\ Cells}$ are the estimates based on $C_{Compatible}$, E_{Old} are the estimates based on C, and Z is the credibility weight assigned to the observations from $C_{Compatible}$. The value of Z may be calculated from the values of E_{Old}, $E_{Compatible\ Cells}$, and E_{New}. Seldom will we be interested in the value of Z if E_{New} is already known. Ultimately, if we know the value of Z, E_{Old}, and $E_{Compatible\ Cells}$, we want to be able to calculate E_{New} via credibility formula (1) above rather than through an additional estimation based on the collective data from C and $C_{Compatible}$. We can derive Z from the statistical assumptions made about the cells. For instance, if we assume the observations from C are from a normally distributed population with mean μ and variance σ^2, the population mean is estimated by the sample mean E_{Old}. If

the separation hypothesis fails, we conclude that the observation from $C_{Compatible}$ are also from a normally distributed population with mean μ and variance σ^2 with the population mean estimated by the sample mean $E_{Compatible\ Cells}$, and the estimate E_{New} of the mean of the population C is given by formula (1) above. The credibility weight Z attached to the mean of the observations from $C_{Compatible}$ is given in Venter [11, p 381] as:

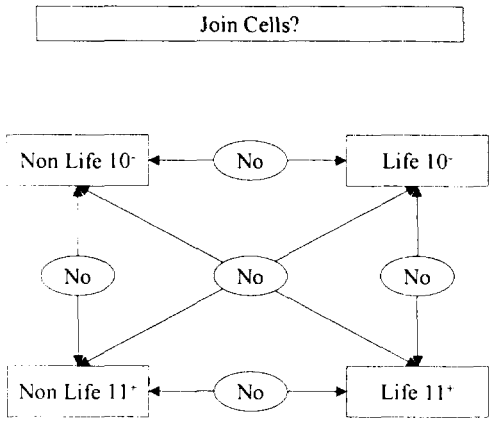
$$Z = n^{-1}/(n^{-1} + m^{-1}) = t^2/(s^2 + t^2) \qquad (2),$$

where m and n represent the number of observations in $C_{Compatible}$ and C, respectively, while s^2 and t^2 are the variances of the means of the observations in $C_{Compatible}$ and C, respectively.

Joining the cells

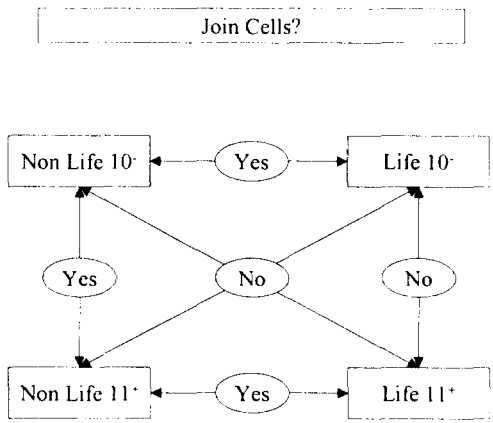
The manner in which cells should be joined for each of the scenarios introduced earlier is shown in figures 5 through 8 below. Under the fourth scenario, the separation hypothesis is true. In that scenario, the parameters underlying the probability model for life actuaries with 10 or fewer years of experience would be estimated based solely on the experience of that cell. The same would apply to the remaining three cells. This is illustrated in figure 5 below:

Figure 5: How to Join Cells in Scenario 4



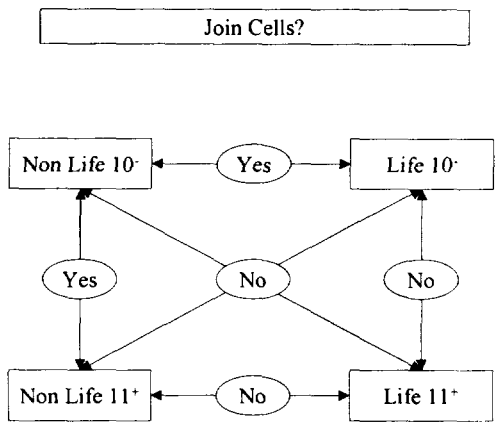
Scenarios one through three represent various alternative hypotheses to the separation hypothesis. In the second scenario, the estimates of the parameters of the models of all four cells would involve other cells as shown in figure 6 below:

Figure 6: How to Join Cells in Scenario 2



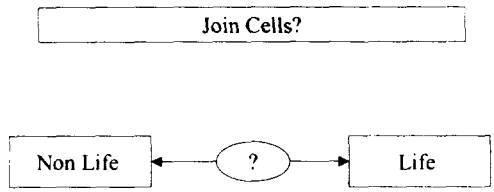
In the third scenario, the cells would be joined as shown in figure 7:

Figure 7: How to Join Cells in Scenario 3



Under the first scenario, all pairs of cells are compatible except for those that are incompatible by definition. To borrow terminology from regression analysis, we may say that the risk characteristics are insignificant and the classification scheme needs to be reconstructed. An alternative classification scheme is provided by dropping one of the rating variables. For instance, by dropping the years of experience variable, we would compare Life versus Non-Life actuaries as shown in figure 8. Alternatively, by dropping the area of practice variable, actuaries with 10 or fewer years of experience would be compared to those with 11or more. If these pairs of new cells are found to be compatible again, then all rating variables are dropped and the original four cells are merged into one to make one set of parameter estimates. If the two new cells are incompatible, then parameters are estimated from each new cell separately.

Figure 8



Defining Risk Classification

Although the definition of various terms that were introduced earlier should be obvious from the context in which they were introduced, let's now attempt to formally define several of these terms.

Risk: "Individual or entity covered by financial security systems [1, p. 2]."

Risk Characteristic: Attribute that identifies a risk or group of risks.

Classification Variable: Categorization or set of risk characteristics consisting of two or more such characteristics. Within a classification variable, risk characteristics define mutually exclusive sets of risks. In other words, a risk can't be identified by more than one characteristic within a classification variable.

Classification Dimension: Number of classification variables used.

Classification Cell: Set of risks sharing all the same risk characteristics.

Adjacent Cells: Two cells C_j and C_k are adjacent if they have exactly D-1 common characteristics where D represents the dimension of the classification scheme.

Non-adjacent Cells: Two cells C_j and C_k are non-adjacent if they have fewer than D-1 common characteristics where D represents the dimension of the Classification scheme.

Cell Universe Ω: Set of all cells defined by the classification variables and risk characteristics.

Classification sample: All observations generated by the risk universe for the process under examination (i.e. frequency, severity). Assume the classification sample is made up of N observations, each observation x_i may be seen as a realization of a random variable X_i where $i = 1,2,...,N$.

Models: Probability Distributions $F_{X_i}(x)$ underlying the random variables in a classification sample. The models underlying the random variables in a classification cell share the same functional form and the same parameters. The parameters of an a priori model for a cell C_j are based on observations from that cell only. The parameters of an a posteriori model for a cell C_j are based on the observations from the class $\{C_j\}$ (see below for a definition of class).

Compatibility: C_k is compatible with C_j if there is a *"reasonable probability" that* the observations in cell C_k could have come from the a priori model (or from a model with the same parameters as) for cell C_j. Technically, the a priori models underlying each cell may have different functional forms. For

instance, the models underlying cells C_j and C_k may be Poisson and Negative Binomial, respectively. For the purpose of assessing compatibility of cell C_k to cell C_j, one asks whether the observations from cell C_k could have come from a Poisson distribution with parameters as in C_j. It is up to the modeler to devise an appropriate test or a set of statistics that can be used to answer the question of compatibility and to define *reasonable probability*. By definition, a cell is compatible with itself.

Incompatibility: C_k is incompatible with C_j if C_k is not compatible with C_j. By definition, we will require that non-adjacent cells be incompatible with one another.

Relation R from Ω to Ω: Non-empty set of ordered pairs (C_j, C_k) such that C_k is compatible with C_j. If $(C_j, C_k) \in R$, we write $C_k R C_j$. If $(C_j, C_k) \notin R$, we write $C_k \not R C_j$. If two cells C_j and C_k are non-adjacent, then by definition $C_k \not R C_j$ and $C_j \not R C_k$

A very important type of relation in set theory is an equivalence relation, which is defined as one having the following three properties:

1) Reflexivity: This property holds that any cell in the cell universe is compatible with itself. We write:

$C_k R C_j$ when $j = k$ or $(C_j, C_j) \in R \; \forall j$.

2) Symmetry: Given any two cells C_j and C_k, if C_j is compatible with C_k then C_k is compatible with C_j and vice versa. We write: $C_k R C_j \Leftrightarrow C_j R C_k$

3) Transitivity: Given any three cells C_j, C_k, and C_l in the cell universe, if C_j is compatible with C_k, and C_k compatible with C_l, it follows that C_j is compatible with C_l. We write:

$C_k R C_j \; and \; C_l R C_k \Rightarrow C_l R C_j$

By definition, the first property always holds for the relation R from Ω to Ω. However, R need be neither symmetric nor transitive. In order words, R need not be an equivalence relation. In Appendix D, we provide an example of an asymmetric relation.

Class $\{C_j\}$: Set of all cells that are compatible with C_j. All cells $C_k \in \Omega$ s.t. $(C_j, C_k) \in R$. Each cell within a classification scheme defines its own class.

Credibility: Weights assigned to the a priori parameter estimates of the cells in a class $\{C_j\}$ in order to come up with the a posteriori parameter estimates for the cell C_j.

Classification Scheme: Process of defining the risks to be covered in a classification scheme, the classification variables and the risk characteristics, the statistical models of the cells, and the rules of compatibility.

Empirical Distribution of the Classification Sample: The empirical distribution of the classification is given by $F_N(x) = \frac{N_x}{N}$ where N_x represents the number of X_k 's such $x_k \leq x$.

Fitted Distribution of the Classification Sample: This distribution is given by $F_Y(x) = \frac{1}{N}\sum_{i=1}^{N} F_{X_i}(x)$, where $F_{X_i}(x)$ represents the a posteriori probability distribution underlying the random variable X_i. If the $F_{X_i}(x)$'s are identical for all the random variables in a cell, then we can write $F_Y(x) = \frac{1}{N}\sum_{j=1}^{n} N_j F_j(x)$, where $F_j(x)$ represents the a posteriori probability distribution underlying the random variables in cell C_j, N_j the number of observations in cell C_j, and n the number of cells in the cell universe.

Illustration

Let's use the classification example presented in our introduction to illustrate our definitions:

Risk: Each actuary represents a risk

Risk Characteristic: Life, Non-Life, 10 or fewer years of experience, 11 or more years of experience.

Classification Variable: Area of practice (life or non-life), Years of experience (10 or fewer, 11 or more).

Classification Dimension: 2.

Classification Cell: For example, life actuaries with 10 or fewer years of experience represent a classification cell.

Adjacent Cells: Two cells are adjacent if they have at least one common characteristic. For instance, Life actuaries with 10 or fewer years of experience and Non-life actuaries with 10 or fewer are adjacent cells.

Non-adjacent Cells: Two cells that have no common characteristics. Life actuaries with 10 or fewer years of experience and Non-life actuaries with 11 or more years of experience are non-adjacent cells.

Cell Universe Ω: Life actuaries with 10 or fewer years of experience, Life actuaries with 11 or more years of experience, Non-Life actuaries with 10 or fewer years of experience, Non-Life actuaries with 11or more years of experience.

Model: $f(x) = \frac{(\lambda d)^x e^{-\lambda d}}{x!}$

Compatibility: $C_k RC_j$ if $\Pr ob(\lambda_j = \lambda_k) \geq .9$.

Let's assume information is collected as per table 1 below:

Table 1

Actuaries	Exposure Units	# of Claims
Life_10	5,000	20
Life_11	10,000	48
Non-Life_10	15,000	88
Non-Life_11	25,000	161

We assume the number of claims in each cell is modeled by a Poisson distribution. The density function of the Poisson distribution is given by: $f(x) = \frac{(\lambda d)^x e^{-\lambda d}}{x!}$ where d is the number of exposure units and λ is the average number of claims per exposure unit. The maximum likelihood estimates $\hat{\lambda}$ of the λ's for each cell of actuary is obtained by dividing the number of claims by the number of exposure units and are shown in table 2 below:

Table 2: MLE Estimates

Actuaries	Exposure Units	# of Claims	MLE Estimate
Life_10	5,000	20	.0040
Life_11	10,000	48	.0048
Non-Life_10	15,000	88	.0059
Non-Life_11	25,000	161	.0064

Recall the two hypotheses introduced in the discussion above.

1) Actuaries within the same cell share the same loss propensity.

2) Actuaries across different cells have different loss propensities.

We will assume that the first hypothesis is true. The second hypothesis can be tested using the following statistic to compare in succession the λ's for pairs of cells C_j and C_k: $\hat{R}_0 = \frac{\hat{\lambda}_j - \hat{\lambda}_k}{\sqrt{\frac{\hat{\lambda}_j}{d_j} + \frac{\hat{\lambda}_k}{d_k}}}$ where $\hat{\lambda}_j$ *and* $\hat{\lambda}_k$ represent the MLE for cells C_j and C_k, respectively, and d_j *and* d_k represent the

exposure units in cells C_j and C_k, respectively. If $\hat{\lambda}_j$ *and* $\hat{\lambda}_k$ are equal (we will refer to the equality of the λ's as a sub-hypothesis) in which case we will say that cells C_j and C_k are compatible, then $\hat{R}_0 \rightarrow N(0,1)$. In other words, $\hat{R}_0$ has the standard normal distribution if cells A and B are compatible. This fact is proven in detail in Appendix A. $\hat{R}_0$ may be thought of as a measure of the distance between the λ's of the two models. For values of $\hat{R}_0$ falling within a given range we will accept the sub-hypothesis that $\hat{\lambda}_j$ *and* $\hat{\lambda}_k$ are equal, while we will reject that sub-hypothesis for values of $\hat{R}_0$ falling outside that range. For instance at a 90% confidence level, the range of real numbers for which we will accept the hypothesis is (-1.645,1.645). We need only calculate $\hat{R}_0$ for adjacent cells. By definition, non-adjacent cells are not compatible. We must reject all the sub-hypotheses in order to accept the main hypothesis. The following two figures 9, 10, and 11show, respectively, the values of $\hat{R}_0$ for the relevant pairs of cells, whether a sub-hypothesis has been accepted or rejected, and whether cells are compatible:

Figure 9: $\hat{R}_0$ values

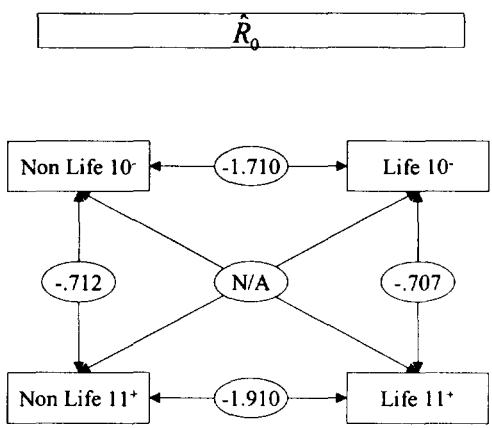


Figure 10: Test Results

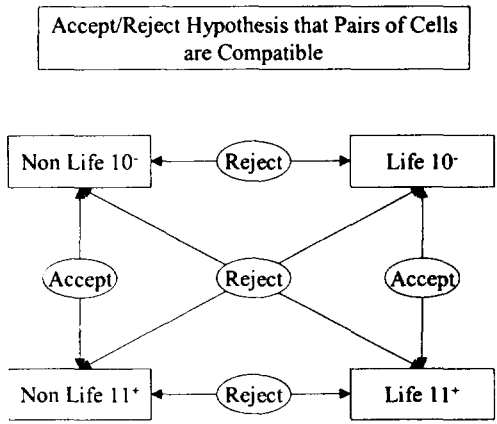
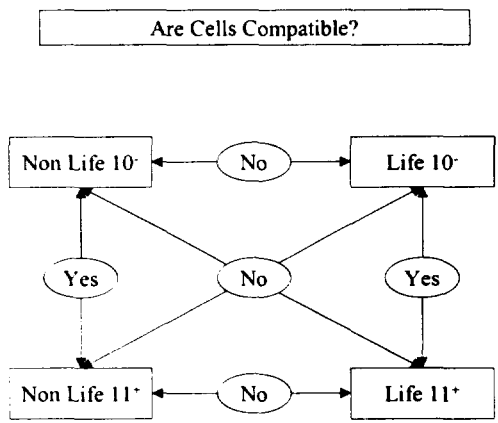


Figure 11: Compatibility Chart



Relation R from Ω to Ω: {(Life_10$^-$,Life_10$^-$), (Non-Life_10$^-$,Non-Life_10$^-$), (Life_11$^+$,Life_11$^+$), (Non-Life_11$^+$,Non-Life_11$^+$) (Life_10$^-$,Life_11$^+$), (Life_11$^+$,Life_10$^-$), (Non-Life_10$^-$,Non-Life_11$^+$), (Non-Life_11$^+$,Non-Life_10$^-$)}

Classes:

{Life_10$^-$} ={Life_10$^-$, Life_11$^+$}

{Life_11$^+$} ={Life_10$^-$, Life_11$^+$}

{Non-Life_10$^-$} ={Non-Life_10$^-$, Non-Life_11$^+$}

{Non-Life_11$^+$} ={Non-Life_10$^-$, Non-Life_11$^+$}

Credibility[7]:

{Life_10$^-$} ={1/3 Life_10$^-$, 2/3 Life_11$^+$}

{Life_11$^+$} ={1/3 Life_10$^-$, 2/3 Life_11$^+$}

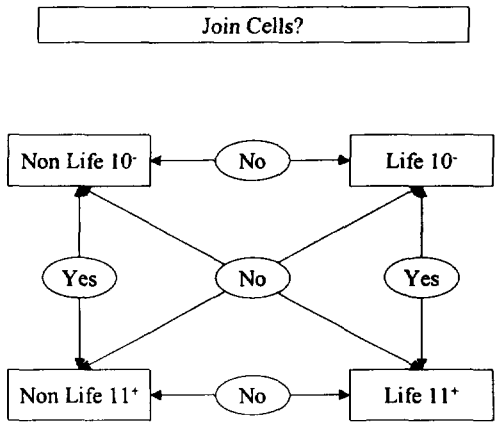
{Non-Life_10$^-$} ={3/8 Non-Life_10$^-$, 5/8 Non-Life_11$^+$}

{Non-Life_11$^+$} ={3/8 Non-Life_10$^-$, 5/8 Non-Life_11$^+$}

Based on figure 11 above, the compatible cells will be joined as shown in figure 12 below in order to produce new estimates of the λ 's. Another way of viewing this is that the new estimates for each cell will be a credibility weighted average of the original estimates of other compatible cells where the weights are given by the relative exposure units of each cell.

[7] For the Poison model the credibility weights for each cell in a class equal the number of exposures in a cell divided by total number of exposures in a class. The derivation is shown in Appendix C.

Figure 12: How to Join Cells



The re-estimated λ 's are as per table 3 below:

Table 3: Revised MLE Estimates

Actuaries	Exposure Units	# of Claims	Initial MLE Estimate	Revised MLE Estimate
Life_10⁻	5,000	20	.0040	.0045
Life_11⁺	10,000	48	.0048	.0045
Non-Life_10⁻	15,000	88	.0059	.0062
Non-Life_11⁺	25,000	161	.0064	.0062

The separation hypothesis has been rejected. The alternative hypothesis that is being accepted here is that both life and non-life actuaries, respectively, have the same loss propensity (expected number of loss per unit of exposure) regardless of their years of experience and that life and non-life actuaries have distinct loss propensities. Hence, the experience of all life actuaries across all years of experience will be combined to arrive at a single estimate of the average claim per exposure unit and the same will be done for non-life actuaries. If, for example, the severity of claims for all actuaries were constant, all life actuaries and all non-life actuaries, respectively, would be charged the same rates. If the separation hypothesis had been accepted, each cell of actuaries would be charged a different rate. In particular, more experienced actuaries

would be charged a higher rate than less experienced ones. The number of claims were simulated from two Poisson distributions for which the actual λ 's are shown in table 4 below. If the main hypothesis had been erroneously accepted, it would lead to subsidies from more experienced actuaries to less experienced ones.

Table 4

Actuaries	Exposure Units	# of Claims	Initial MLE Estimate	Revised MLE Estimate	Actual
Life_10	5,000	20	.0040	.0045	.0050
Life_11*	10,000	48	.0048	.0045	.0050
Non-Life_10	15,000	88	.0059	.0062	.0060
Non-Life_11*	25,000	161	.0064	.0062	.0060

Classification Efficiency

Given a classification scheme, we would like to be able to measure its performance. Classification efficiency is an oft-used notion of performance, which Robert Finger defines as "a measure of a classification system's accuracy [6, p.250]." "A perfect classification system," Finger adds, "would produce the same variability as the insured population. [6, p.250]" Then Finger settles on the squared ratio of the classification system's coefficient of variation (CV) to that of the underlying population as a measure of efficiency. Finger, after observing that, "...the variability [of the insured population] is unknowable," goes on to calculate the efficiency factor for an automobile classification example based on an assumed coefficient of variation of 1.00 for the insured population. This CV of 1.00 is also assumed by Robert Bailey who in his 1960 paper, Any Room Left for Skimming the Cream, uses a similar measure of classification efficiency to Robert Finger's.

The procedure we have outlined makes assumptions about not only the variability but also the actual probability distribution of the underlying population by providing a fitted distribution $F_Y(x)$ to the sample's empirical distribution $F_N(x)$. We could compare the empirical CV of the insured population to that of the fitted distribution. Like Finger and Bailey, we could use some ratio of the CV's as a measure of efficiency. However, the comparison of CV's provides only a limited picture of a classification scheme's

accuracy. We know or have assumed too much about the insured population to rely only on CV ratios to assess the accuracy of the classification scheme. Traditional measures of goodness-of-fit may be more appropriate to evaluate the fit of the assumed distribution vis-à-vis the empirical sample distribution. Two measures immediately come to mind: the Chi Square and the Kolmogorov-Smirnov statistics.

Efficiency, however, should not be thought of as an absolute measure. We would slightly alter Finger's definition of efficiency to read: "efficiency is a measure of a classification system's relative accuracy." What we are in fact measuring is the accuracy of one scheme relative to another. The task becomes one of selecting the classification scheme that best represents the underlying population amongst competing schemes. Each efficiency measure may give a different ranking of the goodness of fit of classification schemes. The modeler may take into account other considerations when making a judgment as to which classification scheme to use.

Validation

Measures of efficiency help us choose the best amongst competing models. However, even the best model might give a poor representation of the data. Validation helps us decide whether the chosen model will be relevant or valid in some future period for which a forecast is sought. If a model fails to validate, we need to rethink the classification scheme and start the process over.

A procedure that could be used to validate a classification scheme consists of randomly selecting out of each cell a percentage of the observations, say 90%, and re-estimate the parameters of the cells through the same process used for the full data set. One then checks to see whether the parameters for each cell fall within an acceptable confidence band of the parameters estimated using the full data set. One may also compare the fitted distribution derived from a 90% sample of the data with that derived from the full data set to see whether the two are "close". This process can be repeated several hundreds or thousands of times using a new random sample each time. A large percentage of the models based on the 90% random samples being consistent with that based on the full data set would tend to validate the original model. One

of the problems with this procedure is that the compatibility of cells will likely depend on the number of observations in the cells. A reduced sample size may affect the compatibility relationships and the composition of the classes. Other validation procedures such as "train and test" and those based on "bootstrap" may be adapted to our classification problem.

When trying to validate a model through the procedure mentioned above, one may need to develop confidence intervals for the original estimates of the parameters of the models so that one could gauge whether the estimates based on the re-sampled data are within an acceptable range of the original estimates. For instance, the standard error of λ in our example is given by $\sqrt{\frac{\hat{\lambda}}{d}}$ and a $k\%$ confidence interval for λ is defined by the interval $\hat{\lambda} \pm z_{(1+k)/2}\sqrt{\frac{\hat{\lambda}}{d}}$, where $z_{(1+k)/2}$ is the $(1+k)/2$ th quantile of the standard normal distribution. The derivation of this interval is shown in Appendix B. The standard error and the 90% confidence interval for λ are shown in table 5 below. A classification scheme that is successfully validated would ensure Predictive Stability, which is one of three actuarial criteria listed by the American Academy of Actuaries in designing a classification scheme.

Table 5: Confidence Interval for λ

Actuaries	Exposure Units	# of Clms	Initial MLE Estimate	Revised MLE Estimate	Std Error of λ	90% Confidence Interval	Actual
Life_10	5,000	20	.0040	.0045	.00055	(.0036,0054)	.0050
Life_11	10,000	48	.0048	.0045	.00055	(.0036,0054)	.0050
Non-Life_10	15,000	88	.0059	.0062	.00039	(.0056,0068)	.0060
Non-Life_11	25,000	161	.0064	.0062	.00039	(.0056,0068)	.0060

Practical Considerations

Earlier in the paper, we stated that separate classification schemes should be used for different aspects of the claim process. The claim process may be decomposed into a frequency and severity component and these components can be further decomposed into more sub-components. We believe that whenever

possible such decomposition may provide a better understanding of the entire claim process. Finally, given how hard it is to find, manipulate, and make inferences about models representing single components of the claim process, the task gets only more daunting when these components are compounded.

Often in insurance problems, there is a need to adjust data for trend and development. Adjustments made to a body of data may cause that data to violate the assumptions of a model. For instance, a Poisson random variable multiplied by a constant is no longer Poisson. If adjustments are made to the data, the model needs to be adjusted accordingly. There may be ways to define the models to see whether any adjustments are appropriate in the first place and the magnitude of such adjustments.

Areas of development

The procedures we have outlined rely on finding good models to represent the probability of random events in a classification cell. There is an extensive library of such models in the literature. In addition, the ability to test the compatibility of cells in a classification scheme is an equally important feature of the procedures presented above. In the illustration, we presented a statistic that allowed us to test the equality of the expected claim per exposure of two Poisson distributions. A number of statistics are available to test hypotheses of the Normal and, by extension, the Lognormal distributions. Various tests and statistics need to be developed in order to make inferences about other distributions, such as the Gamma, Pareto, or the Negative Binomial, that are often used in insurance problems. Distribution of test statistics may also be obtained through simulation rather than heavy-handed calculus.

However, it may not be always feasible to come up with models to represent the probability of events in a cell. Perhaps, there is an even greater role to be played by non-parametric distribution functions and non-parametric approaches to hypothesis testing such as those based on "bootstrap" and "permutation." See Efron and Tibshirani [5] for a discussion of these topics.

Conclusion

The American Academy of Actuaries [2, p. 2] states that the three primary purposes of Risk Classification should be to:

1) protect the insurance program's financial soundness;

2) be fair; and

3) permit economic incentives to operate and thus encourage widespread availability of coverage.

Our definition of Risk Classification is derived out of the very concept of fairness. It is a concept that requires that the same rates not be charged to pools of risks that have fundamentally dissimilar loss propensities or that different rates be charged to pools of risks that have fundamentally similar loss propensities. We believe that the first and third purposes are direct byproducts of the second. The Academy also lists three statistical considerations: homogeneity, credibility and predictive stability. Our definition of credibility differs from that of the Academy. Credibility, as we have defined it, can't be a goal into itself. In lieu of credibility, we would substitute separation as one of the statistical considerations of a classification scheme. If we take this liberty, the purposes and considerations inherent in our definition of risk classification are consistent with those of the American Academy of Actuaries. Our definition provides a definite methodology by which these goals and considerations are met. In addition, nothing in the way we have defined risk classification should preclude us from taking into account other considerations listed by the American Academy of Actuaries including the operational and acceptability considerations.

REFERENCES

[1] Actuarial Standards Board of American Academy of Actuaries, "Actuarial Standard of Practice No. 12, Concerning Risk Classification, (Doc. No. 014)," 1989.

[2] American Academy of Actuaries Committee on Risk Classification, "Risk Classification Statement of Principles," June 1980.

[3] Bailey, Robert A, "Any Room Left for Skimming the Cream?", PCAS XLVII, pp. 30-36.

[4] Casella, Berger, Statistical Inference, Duxbury Press, Belmont, 1990.

[5] Efron,Tibshirani, An Introduction to the Bootstrap, Chapman & Hall / CRC, Boca Raton, 1993/1998.

[6] Finger Robert, Foundations of Casualty Actuarial Science, pp. 231-276, Casualty Actuarial Society, New York, 1990.

[7] Gilbert & Gilbert, Elements of Modern Algebra, Prindle, Weber & Schmidt, Boston, 1984.

[8] Philbrick, Stephen W., "An Examination of Credibility Concepts", PCAS LXVIII, pp.195-219.

[9] Rohatgi, An Introduction to Probability Theory and Mathematical Statistics, John Wiley and Sons, New York, 1976.

[10] Venter, Gary G., "Discussion of minimum Bias with Generalized Linear Models", PCAS LXXVII, pp. 337-349.

[11] Venter, Gary G., Foundations of Casualty Actuarial Science, pp. 375-484, Casualty Actuarial Society, New York, 1990.

[12] Walters, Michael A, "Risk Classification Standards", PCAS LXVIII, pp. 1-18.

APPENDIX A

Let A and B represent two cells, each with frequency distribution defined by a Poisson model with parameters λ_A and λ_B representing the expected number of occurrences per unit of exposure d_i where $d_i \in N$ *and* $1 \le d_i \le d_{\max}$ *for all i's*. Assume that m *and* n observations represented by random variables X_i are made for cells A and B, respectively. Their Poisson models are set as follows:

Random Variable	*Number of Occurrences*	*Exposures Units*	*Mean*	*Variance*
X_1	x_1	d_1	$\lambda_A d_1$	$\lambda_A d_1$
X_2	x_2	d_2	$\lambda_A d_2$	$\lambda_A d_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_m	x_m	d_m	$\lambda_A d_m$	$\lambda_A d_m$
X_{m+1}	x_{m+1}	d_{m+1}	$\lambda_B d_{m+1}$	$\lambda_B d_{m+1}$
X_{m+2}	x_{m+2}	d_{m+2}	$\lambda_B d_{m+2}$	$\lambda_B d_{m+2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{m+n}	x_{m+n}	d_{m+n}	$\lambda_B d_{m+n}$	$\lambda_B d_{m+n}$

The maximum likelihood estimator $\hat{\lambda}_A$ of λ_A is obtained by maximizing the likelihood function

$$L = \frac{\lambda_A^{\sum_{i=1}^{m} x_i} \prod_{i=1}^{m} d_i^{x_i} e^{-\lambda_A \sum_{i=1}^{m} d_i}}{\prod_{i=1}^{m} x_i!}$$

$$\ln(L) = \ln(\lambda_A)\sum_{i=1}^{m} x_i + \sum_{i=1}^{m} x_i \ln(d_i) - \lambda_A \sum_{i=1}^{m} d_i - \sum_{i=1}^{m} \ln(x_i!)$$

$$\frac{d \ln(L)}{d\lambda_A} = \frac{1}{\lambda_A} \sum_{i=1}^{m} x_i - \sum_{i=1}^{m} d_i$$

$$\frac{d\ln(L)}{d\lambda_A} = 0 \Rightarrow \hat{\lambda}_A = \frac{\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} d_i}$$

$$Similarly, \hat{\lambda}_B = \frac{\sum_{i=m+1}^{m+n} x_i}{\sum_{i=m+1}^{m+n} d_i}$$

$\hat{\lambda}_A$ *and* $\hat{\lambda}_B$ are realizations of random variables $\hat{\Lambda}_A = \frac{\sum_{i=1}^{m} X_i}{\sum_{i=1}^{m} d_i}$ *and* $\hat{\Lambda}_B = \frac{\sum_{i=m+1}^{m+n} X_i}{\sum_{i=m+1}^{m+n} d_i}$

$$E(\hat{\Lambda}_A) = \lambda_A \qquad Var(\hat{\Lambda}_A) = \frac{\lambda_A}{\sum_{i=1}^{m} d_i}$$

Also,

$$E(\hat{\Lambda}_B) = \lambda_B \qquad Var(\hat{\Lambda}_B) = \frac{\lambda_B}{\sum_{i=m+1}^{m+n} d_i}$$

Let $\delta = E(\hat{\Lambda}_A) - E(\hat{\Lambda}_B) = \lambda_A - \lambda_B$. Let's define

$$R_\delta = \frac{\hat{\Lambda}_A - \hat{\Lambda}_B - \delta}{\sqrt{\frac{\lambda_A}{\sum_{i=1}^{m} d_i} + \frac{\lambda_B}{\sum_{i=m+1}^{m+n} d_i}}} \text{ and } \hat{R}_\delta = \frac{\hat{\Lambda}_A - \hat{\Lambda}_B - \delta}{\sqrt{\frac{\hat{\Lambda}_A}{\sum_{i=1}^{m} d_i} + \frac{\hat{\Lambda}_B}{\sum_{i=m+1}^{m+n} d_i}}}$$

For $\delta = 0$, we have $R_0 = \frac{\hat{\Lambda}_A - \hat{\Lambda}_B}{\sqrt{\frac{\lambda_A}{\sum_{i=1}^{m} d_i} + \frac{\lambda_B}{\sum_{i=m+1}^{m+n} d_i}}}$ and $\hat{R}_0 = \frac{\hat{\Lambda}_A - \hat{\Lambda}_B}{\sqrt{\frac{\hat{\Lambda}_A}{\sum_{i=1}^{m} d_i} + \frac{\hat{\Lambda}_B}{\sum_{i=m+1}^{m+n} d_i}}}$

Equation 1

We may write $\hat{R}_\delta = \sqrt{\frac{\frac{\lambda_A}{D_A} + \frac{\lambda_B}{D_B}}{\frac{\hat{\Lambda}_A}{D_A} + \frac{\hat{\Lambda}_B}{D_B}}} R_\delta$

Definition 1[4, p. 216]

A sequence of random variables, $X_1, X_2, \ldots$, converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

at all points x where $F_X(x)$ is continuous.

We write $X_n \xrightarrow{L} X$

Definition 2 [4, p. 213]

A sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a random variable X if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{|X_n - X| \geq \varepsilon\} \to 0$$

We write $X_n \xrightarrow{P} X$

We will show that $R_\delta, \hat{R}_\delta, R_0, \hat{R}_0 \xrightarrow{L} Z$, where Z has the standard normal distribution $N(0,1)$.

Then $\hat{R}_0$ can be used to test the null hypothesis

H_0: $\lambda_A = \lambda_B$ versus the alternative

H_1: $\lambda_A \neq \lambda_B$

We accept H_0 at the p confidence level if $|\hat{R}_0| \leq z_{(1+p)/2}$,

and we reject H_0 if $|\hat{R}_0| > z_{(1+p)/2}$.

If the null hypothesis is accepted, we conclude that the claim frequency per unit of exposure underlying cells A and B are equal and we estimate one λ for both cells based on the joint experience of the two. If the null hypothesis is rejected, we say that say that cells A and B have different expected claim frequencies, which are estimated with parameters λ_A and λ_B.

We first show that $R_\delta \xrightarrow{L} Z$.

$$R_\delta = \frac{\hat{\Lambda}_A - \hat{\Lambda}_B - \delta}{\sqrt{\frac{\lambda_A}{\sum_{i=1}^{m} d_i} + \frac{\lambda_B}{\sum_{i=m+1}^{m+n} d_i}}} = \frac{\frac{\sum_{i=1}^{m} X_i}{\sum_{i=1}^{m} d_i} - \frac{\sum_{i=m+1}^{m+n} X_i}{\sum_{i=m+1}^{m+n} d_i} - \lambda_A + \lambda_B}{\sqrt{\frac{\lambda_A}{\sum_{i=1}^{m} d_i} + \frac{\lambda_B}{\sum_{i=m+1}^{m+n} d_i}}}$$

$$R_\delta = \frac{\sum_{i=m+1}^{m+n} d_i \sum_{i=1}^{m} X_i - \sum_{i=1}^{m} d_i \sum_{i=m+1}^{m+n} X_i - \sum_{i=1}^{m} d_i \sum_{i=m+1}^{m+n} d_i \lambda_A + \sum_{i=1}^{m} d_i \sum_{i=m+1}^{m+n} d_i \lambda_B}{\sum_{i=1}^{m} d_i \sum_{i=m+1}^{m+n} d_i \sqrt{\frac{\sum_{i=m+1}^{m+n} d_i \lambda_A + \sum_{i=1}^{m} d_i \lambda_B}{\sum_{i=1}^{m} d_i \sum_{i=m+1}^{m+n} d_i}}}$$

$$R_\delta = \frac{\sum_{i=m+1}^{m+n} d_i \sum_{i=1}^{m} X_i - \sum_{i=1}^{m} d_i \sum_{i=m+1}^{m+n} X_i - \sum_{i=1}^{m} d_i \sum_{i=m+1}^{m+n} d_i \lambda_A + \sum_{i=1}^{m} d_i \sum_{i=m+1}^{m+n} d_i \lambda_B}{\sqrt{\sum_{i=1}^{m} d_i (\sum_{i=m+1}^{m+n} d_i)^2 \lambda_A + \sum_{i=m+1}^{m+n} d_i (\sum_{i=1}^{m} d_i)^2 \lambda_B}}$$

Let $D_A = \sum_{i=1}^{m} d_i$ and $D_B = \sum_{i=m+1}^{m+n} d_i$

We may rewrite $R_\delta = \frac{\sum_{i=1}^{m} D_B X_i - \sum_{i=m+1}^{m+n} D_A X_i - \sum_{i=1}^{m} D_B d_i \lambda_A + \sum_{i=m+1}^{m+n} D_A d_i \lambda_B}{\sqrt{\sum_{i=1}^{m} D_B{}^2 d_i \lambda_A + \sum_{i=m+1}^{m+n} D_A{}^2 d_i \lambda_B}}$

We define:

$$U_i = \begin{matrix} D_B X_i & i = 1,2,\cdots,m \\ -D_A X_i & i = m+1, m+2, \cdots m+n \end{matrix} \quad \text{and} \quad u_i = \begin{matrix} D_B x_i & i = 1,2,\cdots,m \\ -D_A x_i & i = m+1, m+2, \cdots m+n \end{matrix}$$

We then have:

$$\mu_i = E(U_i) = \begin{matrix} D_B d_i \lambda_A & i = 1,2,\cdots,m \\ -D_A d_i \lambda_B & i = m+1, m+2, \cdots, m+n \end{matrix}$$

$$\sigma^2{}_i = Var(U_i) = \begin{matrix} D_B{}^2 d_i \lambda_A & i = 1,2,\cdots,m \\ D_A{}^2 d_i \lambda_B & i = m+1, m+2, \cdots, m+n \end{matrix}$$

We again rewrite $R_\delta = \frac{\sum_{i=1}^{m+n} U_i - \sum_{i=1}^{m+n} \mu_i}{\sqrt{\sum_{i=1}^{m+n} \sigma_i^{\,2}}}$

The Central Limit Theorem - Lindeberg [9, p. 282] provides that the distribution of R_δ converges to the standard normal distribution if the following condition is met:

$$Q = \lim_{m+n \to \infty} \frac{1}{s_{m+n}^2} \sum_{i=1}^{m+n} \sum_{|u_{il} - \mu_i| > \alpha_{m+n}} (u_{il} - \mu_i)^2 p_{il} = 0,$$

where $s_{m+n}^2 = \sum_{i=1}^{m+n} \sigma^2{}_i$ and $p_{il} = \Pr ob(U_i = u_{il})$.

$$p_{il} = \Pr ob(U_i = u_{il}) = \begin{cases} \Pr ob(D_B X_i = u_{il}) \, for \, i = 1,2,...,m \\ \Pr ob(-D_A X_i = u_{il}) \, for \, i = m+1, m+2, ..., m+n \end{cases}$$

$$p_{il} = \begin{cases} \Pr ob(X_i = u_{il} / D_B) \, for \, i = 1,2,...,m \\ \Pr ob(X_i = -u_{il} / D_A \, for \, i = m+1, m+2, ..., m+n \end{cases}$$

$$p_{il} = \Pr ob(X_i = x_{il}) \, where \begin{cases} x_{il} = u_{il} / D_B \, for \, i = 1,2,...,m \\ x_{il} = -u_{il} / D_A \, for \, i = m+1, m+2, ..., m+n \end{cases}$$

$$p_{il} = \begin{cases} \dfrac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} \, for \, i = 1,2,...,m \\ \dfrac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} \, for \, i = m+1, m+2, ..., m+n \end{cases}$$

$$Q = \underset{m+n \to \infty}{Lim} \frac{1}{s_{m+n}^2} \sum_{i=1}^{m+n} \sum_{u_{il} > \varepsilon s_{m+n}} u_{il}^2 \, p_{il+\mu_i} = 0$$

$$p_{il+\mu_i} = \Pr ob(U_i = u_{il} + \mu_i) = \begin{cases} \Pr ob(D_B X_i = u_{il} + \mu_i) \, for \, i = 1,2,...,m \\ \Pr ob(-D_A x_i = u_{il} + \mu_i) \, for \, i = m+1, m+2, ..., m+n \end{cases}$$

$$p_{il+\mu_i} = \Pr ob(X_i = x_{il} + \chi_i) \, where \begin{cases} \chi_i = \mu_i / D_B = E(X_i) \, for \, i = 1,2,...,m \\ \chi_i = -\mu_i / D_A = E(X_i) \, for \, i = m+1, m+2, ..., m+n \end{cases}$$

$$p_{il+\mu_i} = \begin{cases} \dfrac{(\lambda_A d_i)^{x_{il}+\chi_i} e^{-\lambda_A d_i}}{(x_{il}+\chi_i)!} = \dfrac{(\lambda_A d_i)^{\chi_i} x_{il}!}{(x_{il}+\chi_i)!} \dfrac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} \, for \, i = 1,2,...,m \\ \dfrac{(\lambda_B d_i)^{x_{il}+\chi_i} e^{-\lambda_B d_i}}{(x_{il}+\chi_i)!} = \dfrac{(\lambda_B d_i)^{\chi_i} x_{il}!}{(x_{il}+\chi_i)!} \dfrac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} \, for \, i = m+1, m+2, ..., m+n \end{cases}$$

$$p_{il+\mu_i} \le \begin{cases} (\lambda_A d_i)^{\chi_i} \dfrac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} \, for \, i = 1,2,...,m \\ (\lambda_B d_i)^{\chi_i} \dfrac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} \, for \, i = m+1, m+2, ..., m+n \end{cases}$$

$$Q \le \underset{m+n \to \infty}{Lim} \frac{1}{s_{m+n}^2} \left\{ \sum_{i=1}^{m} \sum_{u_{il} > \varepsilon s_{m+n}} u_{il}^2 (\lambda_A d_i)^{\chi_i} \frac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} + \sum_{i=m+1}^{m+n} \sum_{u_{il} > \varepsilon s_{m+n}} u_{il}^2 (\lambda_B d_i)^{\chi_i} \frac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} \right\}$$

$$Q \le \underset{m+n \to \infty}{Lim} \frac{1}{s_{m+n}^2} \left\{ \sum_{i=1}^{m} \sum_{u_{il} > \varepsilon s_{m+n}} \frac{x_{il}^2}{D_B^2} (\lambda_A d_i)^{\chi_i} \frac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} + \sum_{i=m+1}^{m+n} \sum_{u_{il} > \varepsilon s_{m+n}} \frac{x_{il}^2}{D_A^2} (\lambda_B d_i)^{\chi_i} \frac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{S_{m+n}^2} \left\{ \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{\chi_i}}{D_B^2} \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}} x_{il}^2 \frac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{\chi_i}}{D_A^2} \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}} x_{il}^2 \frac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{S_{m+n}^2} \left\{ \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{\chi_i}}{D_B^2} \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}} x_{il} \frac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{(x_{il}-1)!} + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{\chi_i}}{D_A^2} \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}} x_{il} \frac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{(x_{il}-1)!} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{S_{m+n}^2} \left\{ \begin{array}{l} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{\chi_i}}{D_B^2} \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+1} (x_{il}+1) \frac{(\lambda_A d_i)^{x_{il}+1} e^{-\lambda_A d_i}}{x_{il}!} \\ + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{\chi_i}}{D_A^2} \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+1} (x_{il}+1) \frac{(\lambda_B d_i)^{x_{il}+1} e^{-\lambda_B d_i}}{x_{il}!} \end{array} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{S_{m+n}^2} \left\{ \begin{array}{l} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{\chi_i}}{D_B^2} \left[\sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+1} x_{il} \frac{(\lambda_A d_i)^{x_{il}+1} e^{-\lambda_A d_i}}{x_{il}!} + \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+1} \frac{(\lambda_A d_i)^{x_{il}+1} e^{-\lambda_A d_i}}{x_{il}!} \right] \\ + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{\chi_i}}{D_A^2} \left[\sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+1} x_{il} \frac{(\lambda_B d_i)^{x_{il}+1} e^{-\lambda_B d_i}}{x_{il}!} + \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+1} \frac{(\lambda_B d_i)^{x_{il}+1} e^{-\lambda_B d_i}}{x_{il}!} \right] \end{array} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{S_{m+n}^2} \left\{ \begin{array}{l} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{\chi_i}}{D_B^2} \left[\sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+1} \frac{(\lambda_A d_i)^{x_{il}+1} e^{-\lambda_A d_i}}{(x_{il}-1)!} + \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+1} \frac{(\lambda_A d_i)^{x_{il}+1} e^{-\lambda_A d_i}}{x_{il}!} \right] \\ + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{\chi_i}}{D_A^2} \left[\sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+1} \frac{(\lambda_B d_i)^{x_{il}+1} e^{-\lambda_B d_i}}{(x_{il}-1)!} + \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+1} \frac{(\lambda_B d_i)^{x_{il}+1} e^{-\lambda_B d_i}}{x_{il}!} \right] \end{array} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{S_{m+n}^2} \left\{ \begin{array}{l} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{\chi_i}}{D_B^2} \left[\sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+2} \frac{(\lambda_A d_i)^{x_{il}+2} e^{-\lambda_A d_i}}{x_{il}!} + \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+1} \frac{(\lambda_A d_i)^{x_{il}+1} e^{-\lambda_A d_i}}{x_{il}!} \right] \\ + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{\chi_i}}{D_A^2} \left[\sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+2} \frac{(\lambda_B d_i)^{x_{il}+2} e^{-\lambda_B d_i}}{x_{il}!} + \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+1} \frac{(\lambda_B d_i)^{x_{il}+1} e^{-\lambda_B d_i}}{x_{il}!} \right] \end{array} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{s_{m+n}^2} \left\{ \begin{array}{l} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{x_i}}{D_B^2} (\lambda_A d_i)^2 \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+2} \frac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} + (\lambda_A d_i) \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+1} \frac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} \\ + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{x_i}}{D_A^2} (\lambda_B d_i)^2 \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+2} \frac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} + (\lambda_B d_i) \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+1} \frac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} \end{array} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{s_{m+n}^2} \left\{ \begin{array}{l} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{x_i}}{D_B^2} (\lambda_A d_i)^2 \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+2} \frac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} + (\lambda_A d_i) \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_B}+1} \frac{(\lambda_A d_i)^{x_{il}} e^{-\lambda_A d_i}}{x_{il}!} \\ + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{x_i}}{D_A^2} (\lambda_B d_i)^2 \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+2} \frac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} + (\lambda_B d_i) \sum_{|x_{il}|>\frac{\varpi_{m+n}}{D_A}+1} \frac{(\lambda_B d_i)^{x_{il}} e^{-\lambda_B d_i}}{x_{il}!} \end{array} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{s_{m+n}^2} \left\{ \begin{array}{l} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{x_i+2}}{D_B^2} \Pr ob\left[|x_{il}| > \frac{\varpi_{m+n}}{D_B} + 2\right] + \frac{(\lambda_A d_i)^{x_i+1}}{D_B^2} \Pr ob\left[|x_{il}| > \frac{\varpi_{m+n}}{D_B} + 1\right] \\ + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{x_i+2}}{D_A^2} \Pr ob\left[|x_{il}| > \frac{\varpi_{m+n}}{D_A} + 2\right] + \frac{(\lambda_B d_i)^{x_i+1}}{D_A^2} \Pr ob\left[|x_{il}| > \frac{\varpi_{m+n}}{D_A} + 1\right] \end{array} \right\}$$

Using Chebyshev's inequality, we obtain:

$$Q \le \lim_{m+n\to\infty} \frac{1}{s_{m+n}^2} \left\{ \begin{array}{l} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{x_i+2}}{D_B^2} \frac{E(x_{il})}{\frac{\varpi_{m+n}}{D_B}} + \frac{(\lambda_A d_i)^{x_i+1}}{D_B^2} \frac{E(x_{il})}{\frac{\varpi_{m+n}}{D_B}} \\ + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{x_i+2}}{D_A^2} \frac{E(x_{il})}{\frac{\varpi_{m+n}}{D_A}} + \frac{(\lambda_B d_i)^{x_i+1}}{D_A^2} \frac{E(x_{il})}{\frac{\varpi_{m+n}}{D_A}} \end{array} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{s_{m+n}^2} \left\{ \begin{array}{l} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{x_i+2}}{D_B^2} \frac{E(x_{il})}{\frac{\varpi_{m+n}}{D_B}} + \frac{(\lambda_A d_i)^{x_i+1}}{D_B^2} \frac{E(x_{il})}{\frac{\varpi_{m+n}}{D_B}} \\ + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{x_i+2}}{D_A^2} \frac{E(x_{il})}{\frac{\varpi_{m+n}}{D_A}} + \frac{(\lambda_B d_i)^{x_i+1}}{D_A^2} \frac{E(x_{il})}{\frac{\varpi_{m+n}}{D_A}} \end{array} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{s_{m+n}^2} \left\{ \begin{aligned} &\sum_{i=1}^{m} \frac{(\lambda_A d_i)^{\chi_i+2}}{D_B} \frac{E(x_{il})}{\varpi_{m+n}} + \frac{(\lambda_A d_i)^{\chi_i+1}}{D_B} \frac{E(x_{il})}{\varpi_{m+n}} \\ &+ \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{\chi_i+2}}{D_A} \frac{E(x_{il})}{\varpi_{m+n}} + \frac{(\lambda_B d_i)^{\chi_i+1}}{D_A} \frac{E(x_{il})}{\varpi_{m+n}} \end{aligned} \right\}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{s_{m+n}^2} \left\{ \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{\chi_i+3}}{D_B \varpi_{m+n}} + \frac{(\lambda_A d_i)^{\chi_i+2}}{D_B \varpi_{m+n}} + \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{\chi_i+3}}{D_A \varpi_{m+n}} + \frac{(\lambda_B d_i)^{\chi_i+2}}{D_A \varpi_{m+n}} \right\}$$

Recall that $s_{m+n}^2 = \sum_{i=1}^{m} D_B{}^2 d_i \lambda_A + \sum_{i=m+1}^{m+n} D_A{}^2 d_i \lambda_B$

Hence, $s_{m+n}^2 \ge \sum_{i=1}^{m} D_B{}^2 d_i \lambda_A$ *and* $s_{m+n}^2 \ge \sum_{i=m+1}^{m+n} D_A{}^2 d_i \lambda_B$

$$s_{m+n}^2 \ge mn^2 \lambda_A \text{ and } s_{m+n}^2 \ge m^2 n \lambda_B$$

since $1 \le d_i \le d_{\max}$ for all $i = 1,2,\ldots,m+n$,

and $D_A \ge m$ *and* $D_B \ge n$

$$Q \le \lim_{m+n\to\infty} \frac{1}{s_{m+n}^2} \sum_{i=1}^{m} \frac{(\lambda_A d_i)^{\chi_i+3}}{D_B \varpi_{m+n}} + \frac{(\lambda_A d_i)^{\chi_i+2}}{D_B \varpi_{m+n}} + \frac{1}{s_{m+n}^2} \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_i)^{\chi_i+3}}{D_A \varpi_{m+n}} + \frac{(\lambda_B d_i)^{\chi_i+2}}{D_A \varpi_{m+n}}$$

$$Q \le \lim_{m+n\to\infty} \frac{1}{mn^2\lambda_A} \sum_{i=1}^{m} \frac{(\lambda_A d_{\max})^{\chi_i+3}}{n\varepsilon\sqrt{mn^2\lambda_A}} + \frac{(\lambda_A d_{\max})^{\chi_i+2}}{n\varepsilon\sqrt{mn^2\lambda_A}} + \frac{1}{m^2 n\lambda_B} \sum_{i=m+1}^{m+n} \frac{(\lambda_B d_{\max})^{\chi_i+3}}{m\varepsilon\sqrt{m^2 n\lambda_B}} + \frac{(\lambda_B d_{\max})^{\chi_i+2}}{m\varepsilon\sqrt{m^2 n\lambda_B}}$$

Define M such that

$$M \ge (\lambda_A d_{\max})^{\chi_i+3} + (\lambda_A d_{\max})^{\chi_i+2} \text{ and also } M \ge (\lambda_B d_{\max})^{\chi_i+3} + (\lambda_B d_{\max})^{\chi_i+2}$$

We then have, $Q \le \lim_{m+n\to\infty} \frac{M}{n^4 \varepsilon\sqrt{m\lambda_A}} + \frac{M}{m^4 \varepsilon\sqrt{n\lambda_B}} = 0$

Therefore, the Lindeberg condition is met, and $R_\delta \xrightarrow{L} Z$.

We now show that $\hat{R}_\delta \xrightarrow{L} Z$.

From equation 1, we have $\hat{R}_\delta = \sqrt{\frac{\frac{\lambda_A}{D_A} + \frac{\lambda_B}{D_B}}{\frac{\hat{\Lambda}_A}{D_A} + \frac{\hat{\Lambda}_B}{D_B}}} R_\delta$.

The following theorems and statements can be found in or easily verified from Rohatgi [9].

Theorem 1 [9, p.253]

If $X_n \xrightarrow{L} X$ and $Y_n \xrightarrow{P} a$, a constant, then

$$Y_n X_n \xrightarrow{L} aX \qquad \text{if } c \neq 0$$

Theorem 2 [9, p. 245]

Let $X_n \xrightarrow{P} X$ and g be a continuous function defined on $\Re$, then $g(X_n) \xrightarrow{P} g(X)$ as $n \to \infty$.

Corollary 1 [9, p. 245]

$X_n \xrightarrow{P} c$, where c is a constant $\Rightarrow g(X_n) \xrightarrow{P} g(c)$, g being a continuous function.

Statement 1

$X_n \xrightarrow{P} a \Rightarrow \dfrac{a}{X_n} \xrightarrow{P} 1$, where a is a constant.

We first show that $\dfrac{\hat{\Lambda}_A}{D_A} + \dfrac{\hat{\Lambda}_B}{D_B} \xrightarrow{P} \dfrac{\lambda_A}{D_A} + \dfrac{\lambda_B}{D_B}$.

We will show that $K = \underset{m+n\to\infty}{Lim} \Pr ob\left[\left|\dfrac{\hat{\Lambda}_A}{D_A} + \dfrac{\hat{\Lambda}_B}{D_B} - (\dfrac{\lambda_A}{D_A} + \dfrac{\lambda_B}{D_B})\right| > \varepsilon\right] = 0$.

$$\Pr ob\left[\left|\frac{\hat{\Lambda}_A}{D_A} + \frac{\hat{\Lambda}_B}{D_B} - (\frac{\lambda_A}{D_A} + \frac{\lambda_B}{D_B})\right| > \varepsilon\right] = \Pr ob\left[\left|D_B\hat{\Lambda}_A + D_A\hat{\Lambda}_B - D_B\lambda_A - D_A\lambda_B\right| > \varepsilon D_A D_B\right]$$

$$\Pr ob\left[\left|D_B\hat{\Lambda}_A + D_A\hat{\Lambda}_B - D_B\lambda_A - D_A\lambda_B\right| > \varepsilon D_A D_B\right] = \Pr ob\left[\left(D_B\hat{\Lambda}_A + D_A\hat{\Lambda}_B - D_B\lambda_A - D_A\lambda_B\right)^2 > \varepsilon^2 D_A^2 D_B^2\right]$$

Using the Chebyshev inequality, we have

$$K \le \underset{m+n\to\infty}{Lim} \frac{E(D_B\hat{\Lambda}_A + D_A\hat{\Lambda}_B - D_B\lambda_A - D_A\lambda_B)^2}{\varepsilon^2 D_A^2 D_B^2} = \underset{m+n\to\infty}{Lim} \frac{Var(D_B\hat{\Lambda}_A + D_A\hat{\Lambda}_B)}{\varepsilon^2 D_A^2 D_B^2} = \underset{m+n\to\infty}{Lim} \frac{D_B^2\frac{\lambda_A}{D_A} + D_A^2\frac{\lambda_B}{D_B}}{\varepsilon^2 D_A^2 D_B^2}$$

$$K \le Lim \frac{D_B^3\lambda_A + D_A^3\lambda_B}{\varepsilon^2 D_A^3 D_B^3} = \underset{m+n\to\infty}{Lim} \frac{D_B^3\lambda_A}{\varepsilon^2 D_A^3 D_B^3} + \underset{m+n\to\infty}{Lim} \frac{D_A^3\lambda_B}{\varepsilon^2 D_A^3 D_B^3} = \underset{m+n\to\infty}{Lim} \frac{\lambda_A}{D_A^3} + \underset{m+n\to\infty}{Lim} \frac{\lambda_B}{D_B^3} = 0$$

By application of statement 1, we find that

$$\frac{\frac{\lambda_A}{D_A} + \frac{\lambda_B}{D_B}}{\frac{\hat{\lambda}_A}{D_A} + \frac{\hat{\lambda}_B}{D_B}} \xrightarrow{P} 1$$

By corollary 1, we find

$$\sqrt{\frac{\frac{\lambda_A}{D_A} + \frac{\lambda_B}{D_B}}{\frac{\hat{\lambda}_A}{D_A} + \frac{\hat{\lambda}_B}{D_B}}} \xrightarrow{P} 1$$

and finally, by application of theorem 1,

$$\hat{R}_S = \sqrt{\frac{\frac{\lambda_A}{D_A} + \frac{\lambda_B}{D_B}}{\frac{\hat{\lambda}_A}{D_A} + \frac{\hat{\lambda}_B}{D_B}}} R_S \xrightarrow{L} N(0,1)$$

The proof is complete.

APPENDIX B

Confidence Interval for λ

Let $\hat{\lambda}$ be the MLE of λ for Poisson distributed random variables X_i, $i = 1,2,...,n$, with means λd_i and variances λd_i where d_i is the number of exposure units associated with X_i. Let x_i, $i = 1,2,...,n$ be the realization of the random variables X_i.

Then, $\hat{\lambda} = \dfrac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} d_i}$ is the realization of a random variable $\hat{\Lambda}$, where $\hat{\Lambda} = \dfrac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} d_i}$.

$\hat{g} = \dfrac{\hat{\lambda} - \lambda}{\sqrt{\dfrac{\hat{\lambda}}{\sum_{i=1}^{N} d_i}}}$ is the realization of the random variable $\hat{G} = \dfrac{\hat{\Lambda} - \lambda}{\sqrt{\dfrac{\hat{\Lambda}}{\sum_{i=1}^{N} x_i}}}$

Using the definition introduced in appendix A, we will show $\hat{G} \xrightarrow{L} Z$ where Z has the standard normal distribution, and a $k\%$ confidence interval for λ is $\hat{\lambda} \pm z_{(1+k)/2} \sqrt{\dfrac{\hat{\lambda}}{\sum_{i=1}^{N} d_i}}$ where $z_{(1+k)/2}$ is the

$(1+k)/2$ *th quantile of the standard normal distribution.*

Let's now prove that $\hat{G} \xrightarrow{L} Z$.

Let $G = \dfrac{\hat{\Lambda} - \lambda}{\sqrt{\dfrac{\lambda}{\sum_{i=1}^{N} x_i}}}$. Observe that $\hat{G} = \sqrt{\dfrac{\lambda}{\hat{\lambda}}} \dfrac{\hat{\lambda} - \lambda}{\sqrt{\dfrac{\lambda}{\sum_{i=1}^{N} d_i}}} = \sqrt{\dfrac{\lambda}{\hat{\lambda}}} G$

We first prove $G \xrightarrow{L} Z$

We rewrite $G = \dfrac{\dfrac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} d_i} - \lambda}{\sqrt{\dfrac{\lambda}{\sum_{i=1}^{N} d_i}}} = \dfrac{\sum_{i=1}^{N} X_i - \lambda d_i}{\sqrt{\sum_{i=1}^{N} \lambda d_i}} = \dfrac{\sum_{i=1}^{N} X_i - E(X_i)}{\sqrt{\sum_{i=1}^{N} Var(X_i)}}$

The Central Limit Theorem- Linderberg [9, p.282] states that the distribution of G converges to the standard normal distribution provided that the following condition is met:

$$L = \lim_{n\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} \sum_{|x_{il}-\mu_i|>\varepsilon s_n} (x_{il} - \mu_i)^2 p_{il} = 0,$$

where $\mu_i = E(X_i), s_n^2 = \sum_{i=1}^{n} Var(X_i), and\ p_{il} = \Pr ob(X_i = x_{il})$.

$$L = \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} \sum_{|x_{il}|>\varepsilon s_N} x_{il}^2 p_{il+\mu_i}$$

$$p_{il+\mu_i} = \frac{(\lambda d_i)^{x_{il}+\mu_i} e^{-\lambda d_i}}{(x_{il}+\mu_i)!} = \frac{(\lambda d_i)^{\mu_i} x_{il}!}{(x_{il}+\mu_i)!} \frac{(\lambda d_i)^{x_{il}} e^{-\lambda d_i}}{x_{il}!} \le (\lambda d_i)^{\mu_i} \frac{(\lambda d_i)^{x_{il}} e^{-\lambda d_i}}{x_{il}!}$$

$$L \le \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} \sum_{|x_{il}|>\varepsilon s_N} x_{il}^2 (\lambda d_i)^{\mu_i} \frac{(\lambda d_i)^{x_{il}} e^{-\lambda d_i}}{x_{il}!} = \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} (\lambda d_i)^{\mu_i} \sum_{|x_{il}|>\varepsilon s_N} x_{il}^2 \frac{(\lambda d_i)^{x_{il}} e^{-\lambda d_i}}{x_{il}!}$$

$$L \le \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} (\lambda d_i)^{\mu_i} \sum_{|x_{il}|>\varepsilon s_N} x_{il} \frac{(\lambda d_i)^{x_{il}} e^{-\lambda d_i}}{(x_{il}-1)!} = \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} (\lambda d_i)^{\mu_i} \sum_{|x_{il}|>\varepsilon s_N+1} (x_{il}+1) \frac{(\lambda d_i)^{x_{il}+1} e^{-\lambda d_i}}{x_{il}!}$$

$$L \le \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} (\lambda d_i)^{\mu_i} \left[\sum_{|x_{il}|>\varepsilon s_N+1} x_{il} \frac{(\lambda d_i)^{x_{il}+1} e^{-\lambda d_i}}{x_{il}!} + \sum_{|x_{il}|>\varepsilon s_N+1} \frac{(\lambda d_i)^{x_{il}+1} e^{-\lambda d_i}}{x_{il}!} \right]$$

$$L \le \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} (\lambda d_i)^{\mu_i} \left[\sum_{|x_{il}|>\varepsilon s_N+1} \frac{(\lambda d_i)^{x_{il}+1} e^{-\lambda d_i}}{(x_{il}-1)!} + \sum_{|x_{il}|>\varepsilon s_N+1} \frac{(\lambda d_i)^{x_{il}+1} e^{-\lambda d_i}}{x_{il}!} \right]$$

$$L \le \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} (\lambda d_i)^{\mu_i} \left[\sum_{|x_{il}|>\varepsilon s_N+2} \frac{(\lambda d_i)^{x_{il}+2} e^{-\lambda d_i}}{x_{il}!} + \sum_{|x_{il}|>\varepsilon s_N+1} \frac{(\lambda d_i)^{x_{il}+1} e^{-\lambda d_i}}{x_{il}!} \right]$$

$$L \le \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} (\lambda d_i)^{\mu_i} \left[(\lambda d_i)^2 \sum_{|x_{il}|>\varepsilon s_N+2} \frac{(\lambda d_i)^{x_{il}} e^{-\lambda d_i}}{x_{il}!} + (\lambda d_i) \sum_{|x_{il}|>\varepsilon s_N+1} \frac{(\lambda d_i)^{x_{il}} e^{-\lambda d_i}}{x_{il}!} \right]$$

$$L \le \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} (\lambda d_i)^{\mu_i+2} \Pr ob\left[|x_{il}| > \varepsilon s_N + 2\right] + (\lambda d_i)^{\mu_i+1} \Pr ob\left[|x_{il}| > \varepsilon s_N + 1\right]$$

Using Chebyshev's inequality, we obtain:

$$L \le \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} (\lambda d_i)^{\mu_i+2} \frac{E(x_{il})}{\varepsilon s_N + 2} + (\lambda d_i)^{\mu_i+1} \frac{E(x_{il})}{\varepsilon s_N + 1} = \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} \frac{(\lambda d_i)^{\mu_i+3}}{\varepsilon s_N + 2} + \frac{(\lambda d_i)^{\mu_i+2}}{\varepsilon s_N + 1}$$

$$L \le \lim_{N\to\infty} \frac{1}{s_N^2} \sum_{i=1}^{N} \frac{(\lambda d_i)^{\mu_i+3} + (\lambda d_i)^{\mu_i+2}}{\varepsilon s_N}$$

Since $1 \le d_i \le d_{\max}$ for all $i = 1,2,...,N$.

$$s_N^2 = \sum_{i=1}^{N} d_i \lambda \ge N\lambda \quad \text{and } (\lambda d_i)^{\mu_i+3} + (\lambda d_i)^{\mu_i+2} \le M$$

Therefore, $L \le \lim_{N\to\infty} \frac{1}{N\lambda} \sum_{i=1}^{N} \frac{M}{\varepsilon\sqrt{N\lambda}} = \lim_{N\to\infty} \frac{1}{\lambda} \frac{M}{\varepsilon\sqrt{N\lambda}} = 0$

Hence, the Linderberg condition is satisfied and $G \xrightarrow{L} Z$.

We now show that $\hat{G} \xrightarrow{L} Z$.

$$\hat{G} = \sqrt{\frac{\lambda}{\hat{\lambda}}} G$$

We first show $\hat{\lambda} \xrightarrow{P} \lambda$.

We will show that $K = \underset{N \to \infty}{Lim} \Pr ob\left[\left|\hat{\lambda} - \lambda\right| > \varepsilon\right] = 0$

Using Chebyshev's inequality,

$$K \le \underset{N \to \infty}{Lim} \frac{E(\hat{\lambda} - \lambda)^2}{\varepsilon^2} = \underset{N \to \infty}{Lim} \frac{Var(\hat{\lambda})}{\varepsilon^2} = \underset{N \to \infty}{Lim} \frac{\lambda}{\varepsilon^2 \sum_{i=1}^{N} d_i} \le \underset{N \to \infty}{Lim} \frac{\lambda}{\varepsilon^2 N} = 0.$$

By application of statement 1 of Appendix A, we find

$$\frac{\lambda}{\hat{\lambda}} \to 1$$

By application of corollary 1,

$$\sqrt{\frac{\lambda}{\hat{\lambda}}} \to 1$$

Finally, by application of theorem 1

$$\hat{G} = \sqrt{\frac{\lambda}{\hat{\lambda}}} G \to N(0,1).$$

Our proof is complete.

APPENDIX C

Assume there are n cells in a class, and the MLE estimate $\hat{\lambda}_j$ of cell C_j, for $j = 1,2,...n$, is given by

$\hat{\lambda}_j = \dfrac{\sum_{i=1}^{N_j} x_i}{\sum_{i=1}^{N} d_i}$ where N_j is the number of observations in cell C_j. The MLE estimate for the class is given by $\hat{\lambda} = \dfrac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} d_i}$ where N is the total number of observations across all cells such that $\sum_{j=1}^{n} N_j = N$.

We rewrite $\hat{\lambda}$ as:

$$\hat{\lambda} = \frac{\sum_{i=1}^{N_1} x_i + \sum_{i=N_1+1}^{N_2} x_i + \ldots + \sum_{i=N_{n-1}+1}^{N_n} x_i}{\sum_{i=1}^{N_1} d_i + \sum_{i=N_1+1}^{N_2} d_i + \ldots + \sum_{i=N_{n-1}+1}^{N_n} d_i}$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{N_1} x_i}{\sum_{i=1}^{N} d_i} \frac{\sum_{i=1}^{N_1} d_i}{\sum_{i=1}^{N_1} d_i} + \frac{\sum_{i=N_1+1}^{N_2} x_i}{\sum_{i=1}^{N} d_i} \frac{\sum_{i=N_1+1}^{N_2} d_i}{\sum_{i=N_1+1}^{N_2} d_i} + \ldots + \frac{\sum_{i=N_{n-1}+1}^{N_n} x_i}{\sum_{i=1}^{N} d_i} \frac{\sum_{i=N_{n-1}+1}^{N_n} d_i}{\sum_{i=N_{n-1}+1}^{N_n} d_i}$$

$$\hat{\lambda} = \hat{\lambda}_1 \frac{\sum_{i=1}^{N_1} d_i}{\sum_{i=1}^{N} d_i} + \hat{\lambda}_2 \frac{\sum_{i=N_1+1}^{N_2} d_i}{\sum_{i=1}^{N} d_i} + \ldots + \hat{\lambda}_n \frac{\sum_{i=N_{n-1}+1}^{N_n} d_i}{\sum_{i=1}^{N} d_i}$$

APPENDIX D

Assume two cells C_1 and C_2 with the following ten observations:

C_1	C_2
10.154	8.508
11.510	8.100
5.453	11.707
13.239	8.772
10.065	14.156
(2.307)	6.953
17.625	7.612
13.242	10.633
14.319	7.463
7.619	5.546

The observations in C_1 are assumed to come from a Normal distribution with cumulative probability function F_1, while those in C_2 are assumed to come from a Lognormal distribution with cumulative probability function F_2.

Compatibility is defined as follows:

> Given two cells C_i and C_j, C_i is compatible to C_j if: $F_j(x_{ik}) > 0$ *for all* $k = 1,2,\ldots n$, where x_{ik} is an observation from C_i, n the number of observations in C_i, and F_j the cumulative probability function for cell C_j.

Since $F_1(x_{2k}) > 0$ *for all* $k = 1,2,\ldots 10$, we say that C_2 is compatible to C_1. However, $F_2(-2.307) = 0$, therefore we say that C_1 is not compatible to C_2.