

Empirical Bayes Methods in Road Safety Research

D-97-13

Dr. Rian A.W. Vogelesang

Leidschendam, 1997

SWOV Institute for Road Safety Research, The Netherlands

Report documentation

Number: D-97-13
Title: Empirical Bayes Methods in Road Safety Research
Author(s): Dr. Rian A.W. Vogelesang
Research manager: S. Oppe

Keywords: Accident, calculation, mathematical model, regression analysis .
Contents of the project: This report gives an overview of the most influential paradigms and techniques in road safety research. Discussed are analytical tools which go under names as 'prior' or 'posterior', 'overdispersed Poisson', 'negative binominal', 'Gamma', 'negative regression effect' 'accident proneness vs liability' etc.

Number of pages: 44 p.
Price: Dfl. 22,50
Published by: SWOV, Leidschendam, 1997

SWOV Institute for Road Safety Research
P.O. Box 1090
2260 BB Leidschendam
The Netherlands
Telephone 31703209323
Telefax 31703201261

Summary

Road safety research is a wonderful combination of counting fatal accidents and using a toolkit containing prior, posterior, overdispersed Poisson, negative binomial and Gamma distributions, together with positive and negative regression effects, shrinkage estimators and fierce debates concerning the phenomenon of accident migration. Accidents are counted at the level of, e.g., roundabouts of some specific architecture and also *over* all roundabouts. For any *individual* roundabout a Poisson distribution is used to assess its 'unsafety'; for the *mean* of the accidents over roundabouts of the same architecture, a Gamma distribution is used. The combination of the two leads to a *mixed* distribution, 'a negative binomial distribution' for the accidents for all roundabouts together. Methodological problems are the regression effect, the *apparent* reduction of fatal accidents after remedial treatment, the selection of, e.g., roundabouts, for remedial treatment, and assessing the influence of the remedial treatment. For the selection of systems for remedial treatment, decision theoretical concepts like 'false' and 'correct' positives and negatives are used.

Contents

1.	<i>Introduction</i>	6
1.1.	Estimation of 'Unsafety'	7
1.2.	Effect of Remedial Treatment	8
1.3.	Regression to the Mean	8
1.4.	Expected Number of Accidents	9
1.5.	Migration of Accidents	9
1.6.	Organisation of the Subjects	10
2.	<i>Accident Proneness, Compound Poisson & Mixtures</i>	11
2.1.	History of Accident Statistics	11
2.2.	Accident Proneness	11
2.3.	Compound Poisson: Mixture of Distributions	13
2.3.1.	Two Sub-Periods: Bivariate Compound Poisson Distribution	13
2.3.2.	Estimating the (Bivariate) Negative Binomial Distribution	14
3.	<i>Shrinkage Estimators: Stein's Paradox</i>	15
4.	<i>Accident Distributions</i>	16
4.1.	Distributions: Poisson, Gamma, and Negative Binomial	16
5.	<i>Classical vs Empirical Bayes: Robbins (1956)</i>	17
5.1.	Λ Stochastic	18
5.2.	Mixtures: Mixed Poisson Distribution	18
6.	<i>Estimation for Mixtures of the Poisson Distribution</i>	19
6.1.	Estimation for the Compound Poisson: G Unknown	19
6.2.	Restricted Empirical Bayes Method	20
6.3.	Empirical Distribution G Completely Specified	21
6.4.	Estimation for Compound Poisson: Various Statistics	22
7.	<i>Prediction for Mixtures of the Exponential Distribution (Including the Compound Poisson)</i>	23
7.1.	$E(\lambda X > a)$: Mixture of Exponential Distributions	24
8.	<i>Regression to the Mean: Abbess, Jarrett & Wright</i>	25
8.1.	The Regression Effect	27
9.	<i>Linear Empirical Bayes (LEB) Estimation</i>	28
9.1.	Two-Period Regression: 'Before' and 'After'	29
9.2.	Computing the Regression Effect	29
10.	<i>Estimating Safety</i>	30
10.1.	Poisson-Gamma Model	31
10.2.	Parameter Estimation	31
11.	<i>Bias by Selection: Hauer, Hauer & Persaud</i>	32
11.1.	'Before' and 'After' Period are Equal	32
11.1.1.	Size of Selection Bias	33
11.1.2.	Selection Bias is One-Sided	33
11.2.	The Shaman	33
11.2.1.	Expected Number of Accidents	34
11.3.	Empirical Bayes vs Nonparametric	34
11.3.1.	Nonparametric Method	35
11.3.2.	Empirical Bayes Method	35
11.3.3.	Comparison of Both Methods	35
11.4.	The Sieve	35
11.4.1.	Sieve Efficiency	36
12.	<i>Migration of Accidents</i>	37
13.	<i>Overview of Types of Distributions and Statistics</i>	39
14.	<i>References</i>	40

1. Introduction

This report is an account of research at SWOV Institute for Road Safety Research in the field of road safety research, inspired by papers like Abbess, Jarrett & Wright (1981), Jarrett, Abbess & Wright (1982), Hauer (1980, 1986, 1992, 1995), Hauer & Persaud (1983, 1984, 1987), Robbins (1950, 1951, 1956, 1964, 1977, 1980*a,b,c*) - among many others - and starting with Arbous & Kerrich's (1951) influential paper in Biometrics. This research is a wonderful combination of practice in the sense of counting fatal accidents in e.g., the county of Hertfordshire, and a typical set of analytical tools which go under names as 'prior' or 'posterior', 'overdispersed Poisson', 'negative binomial', 'Gamma', 'negative regression effect', 'accident proneness vs liability', etc. Although many papers have been published in various scientific periodicals on methodological issues in road safety research, I couldn't find a comprehensive book or overview on this subject. The aim of this paper is to give an overview of the most influential paradigms and techniques in road safety research. This may explain the selection of the papers for discussion in this study. In the following, we first give an overview of some general issues like estimating 'unsafety', the effect of remedial treatment and regression to the mean, the expected number of accidents and accident migration.

Central topics in road safety research with related statistical characteristics, are e.g., the estimation of 'unsafety', the expected frequency of accidents at road sections, at intersections, or of drivers, etc.; identification of unsafe traffic entities; estimating the effect of remedial treatment of unsafe traffic systems, including the possibility of migration of accidents, and prediction of the frequency of future accidents. Apart from the unsafety due to road characteristics, at least some 'unsafety' is due to personal factors. There is a difference between accident *proneness* (depending on psycho-physiological qualities, such as health, age, fatigue, experience, use of alcohol and drugs) and accident *liability* (due to environmental factors, e.g., exposition: number of kilometers per year). The accident-prone person is regarded as one who has many accidents: at home, at work, on the public highways, and is in fact a sort of 'calamity Joe' (Arbous & Kerrich, 1951), although others deny the existence of such a person (Koomstra, 1978; Asmussen & Kranenburg, 1985).

The hypothesis of 'chance distribution' assumes that accidents occur to individuals by pure chance, leading to the Poisson distribution for the distribution of accidents. Although numbers of accidents in two subsequent periods each may follow the Poisson distribution, they may not follow a Poisson distribution when taken together: they may follow a negative binomial distribution. The hypothesis of unequal liabilities results in groups being nonhomogeneous in their sustainability of accidents, and leads to the negative binomial distribution instead of the Poisson distribution.

Different hypotheses use different forms of the negative binomial distribution, which has become a standard for the analysis of accident data. Many of the related problems regard sets of Poisson distributed variables with different Poisson parameters (λ). Both in pure Bayes and in Empirical Bayes (EB) estimation, the parameter value (λ) itself is regarded as a realisation of a random variable Λ having a Gamma distribution, $G(\lambda)$, the prior distribution. In the pure Bayes approach, $G(\lambda)$ is not necessarily interpreted in terms of relative frequencies, but in the EB approach it is given a frequency interpretation (Robbins, 1956). Also, the availability of previous data, suitable of estimation of $G(\lambda)$, is assumed. The EB estimation rule generally depends on all past and current observed accident frequencies (X).

1.1. Estimation of 'Unsafety'

'Unsafety' is an attribute of a specific road section, intersection, a driver, etc. The estimation of unsafety is by assessing numbers of accidents (corrected for intensity of traffic) and their negative consequences. Counting numbers of fatalities is an estimate of unsafety, but often not an unbiased one, because of the effect of 'regression to the mean'. For example, when selecting a roundabout that had a very bad accident record last year, we expect, on average, less accidents for that roundabout next year, if nothing changes.

Numbers of accidents x at a fixed location are often assumed to be Poisson distributed. These sites are combined into a group, but within the group, the average numbers λ for the separate sites may show variation (see above). The effectiveness of remedial measures is based on the comparison between accident frequencies before and after treatment. However, this comparison is complicated by a number of factors, such as regression-to-the-mean and, possibly, accident migration (Boyle & Wright, 1984, 1985; Persaud, 1984; Maher, 1987, 1990). Because accident frequencies at distinct sites within a group do vary, the distribution of accidents at these sites is not exactly Poisson distributed, but may be better characterised by an 'overdispersed' or 'compound' Poisson distribution (Arbous et al., 1951; Robbins, 1951, 1956). If the Poisson parameters follow a Gamma distribution, the accidents within the groups follow a negative binomial distribution.

The unknown accident rate at a blackspot is regarded as the value of a random variable having a probability distribution (Jarrett et al., 1982). This distribution is regarded as the 'prior distribution' for this blackspot. Given the known accident frequency over a recent period, this 'prior' distribution is converted into a 'posterior' distribution using Bayes' theorem. If the prior distribution is a Gamma distribution, and if the number of accidents is Poisson distributed, then the posterior distribution is again a Gamma distribution (Robbins, 1956; Gipps, 1980; Abbess et al., 1981; Jarrett et al., 1982). The parameters of the posterior distribution, $f(\lambda | x)$, are easily derived from the prior distribution, $f(\lambda)$, and the observed data, $f(x | \lambda)$. The mean of the posterior distribution, $E(\lambda | x)$, is the best predictor of the accident rate in the next year (given that the true accident rate λ remains constant). The size of the regression effect is equal to the difference between $E(\lambda | x)$ and x (see Figure 1). Let α and β denote the length of the observation period and the number of fatalities, respectively. The regression line is a linear function of X , with slope $1/(1 + \alpha)$ (Jarrett et al., 1982[†]). The regression line intersects the 45° line where $X = \beta/\alpha$. The regression effect is negative for values of X larger than β/α , indicating an expected *reduction* in the number of accidents.

The mean of the posterior distribution of Λ for some X , $E(\lambda | x)$, is the EB point estimate of the true accident frequency (Arbous et al., 1951; Robbins, 1956). The EB estimate is corrected for the effect of 'regression to the mean' (Abbess et al., 1981; Jarrett et al., 1982, 1988 a.o.), by conditioning on x . This is illustrated in Figure 1. The random variable X (X -axis) denotes observed numbers of accidents. The Y -axis represents $E(\lambda | x)$: the expected accident frequency for a given value x . When $Y = X = \beta/\alpha$, the expected value of Y is on the regression line, 'the line of expectations' (Freedman, Pisani & Purves, 1978, p. 161; Jarrett et al., 1982). The effectiveness of remedial measures is based on a comparison between accident frequencies before and after treatment.

[†]Jarrett et al. (1982) use s_0 and n_0 instead of α and β .

1.2. Effect of Remedial Treatment

Bayesian analysis differs significantly from the classical analysis of accident data. The motivation for the Bayesian approach is the desire to use prior knowledge about accidents in the analysis. This prior knowledge is often based on regional characteristics and accident history at that location, and is used to determine the extent to which the site is hazardous. In a first step, accident histories are aggregated over sites (or years) to estimate the probability distribution of accident rates. In the second step this (empirical) prior distribution and the accident history at a particular site - which is assumed to be Poisson - are used to obtain an estimate of the (posterior) mixed probability distribution at the particular site. For this, nonparametric methods and Bayesian methods are available (e.g., Hauer, 1980a,b, 1981, 1982, 1986, 1992, 1995). Maher & Mountain (1988) discuss two methods ('annual accident total' vs 'potential accident reduction') for the identification of potential blackspots and the evaluation of the treatment, including the regression-to-mean effect. Outstanding results are those by Robbins, Hauer, Hauer & Persaud, and Abbess, Jarrett & Wright.

1.3. Regression to the Mean

One of the problems is the selection of entities for improvement. Notoriously dangerous entities are, on average, less dangerous after remedial treatment, but some entities would have 'improved' as well without remedial treatment. On the grounds of the principle of 'regression to the mean', we expect for these entities lower accident frequencies in the future. *Because* they have the higher accident rates, they are selected for improvement. This is the meaning of 'bias by selection', according to Hauer (1980a). For these sites, a decrease in the number of accidents is expected, *irrespective of* the success of the remedial treatment. In the same way, we expect an increase in the number of accidents for systems that have less than average accidents: the so-called *positive* regression effect (Jarrett et al, 1982; McGuigan, 1985b). The regression to the mean effect results from *selecting* objects for investigation. If, e.g., car drivers are selected on the basis of their previous accident records, then the average tends to decrease for the group with accidents and to increase for the group without accidents, for purely statistical reasons. When the duration of the observation period increases, the relative size of the regression effect diminishes. This applies to drivers as well: drivers with one or more violations in one period are seen to have, on average, 60-80 percent less violations in the next period (Hauer et al., 1983). The effect is much larger for drivers than for road segments (because of hospitalisation, license revocation, etc.). A problem is how to obtain unbiased estimates in before-and-after studies in the presence of the regression effect (but see Jarrett, 1994). All authors assume that the number of accidents on each system follows a Poisson or a compound Poisson distribution (Robbins, 1964, 1977, 1980a,b,c; Hauer, 1980; Gipps, 1980; Abbess et al., 1981, a.o.).

Most empirical Bayes estimators are shrinkage estimators, and like the LEB ('Linear Empirical Bayes') estimator, they 'shrink' the observed accident frequency towards the *grand average*, the expected value (Stein, 1955; Efron & Morris, 1977; Jarrett et al., 1982, 1988; Jarrett, 1991, 1994). This results in a smaller mean squared error of estimate. The size of the regression-to-the-mean effect and its dependence on length of observation period is clearly described by Abbess et al., 1981, and by Wright et al., 1988. Identification of blackspots is described in a decision theoretic framework by Hauer & Persaud (1984). Notoriously dangerous systems are, on average, less dangerous after remedial treatment, but some systems would have 'improved' as well without remedial treatment (see section 4.4). In the words of Hauer & Persaud (p. 165):

"Thus, prayers for accident reduction on road sections that in one year had three or more accidents (...) might seem to bring about a 55 percent reduction in the number of accidents."

From the principle of 'regression to the mean', we expect for these systems lower accident frequencies in the future. But because they have the higher accident rates, they are selected for improvement. This is the meaning of 'bias by selection', according to Hauer (1980a). In short, we expect a decrease in the number of accidents *irrespective of the success of the remedial treatment for systems*. In the same way, we expect an increase in the number of accidents for systems that have less than average accidents: the so-called *positive regression effect* (Jarrett et al, 1982; McGuigan, 1985b).

1.4. Expected Number of Accidents

If systems are grouped on the basis of their accident history, the average number of accidents \bar{x} observed for a group of systems (crossings, roundabouts, drivers, etc.) in some period need not be equal to the number of accidents in the subsequent period (Hauer, 1986). There are large discrepancies between x (as an estimator of the average, λ), and a , the number of accidents in the after period (as an estimate of the average of the λ 's of, for example, roundabouts with equal x).

Were x a good estimator for λ , then systems which recorded x accidents in one period would record, on average, x accidents in the next period of equal duration if their expected number of accidents did not change. However, this is not generally observed. Therefore, the conclusion must be that x is not a good estimate of λ (Hauer, 1982, 1986). Hauer (1986) even makes a stronger claim: x is not a good estimate of λ irrespective of how systems are selected. (This may be because x is small and, hence, may be unreliable).

Nevertheless, many studies still use the 'before-and-after' design or 'with-without' studies. Even random selection for treatment cannot eliminate the bias (Jarrett, 1988, p. 94). This is because of the difficulty of obtaining two exactly similar sites. The treated and untreated sites should be chosen in exactly the same way, they should have similar accident frequencies in the before period. Also, the site to be treated should then be chosen at random, by the toss of a coin. A significant result is unlikely to be obtained because of the small frequencies at both the treated and control sites. The conclusion is that one just should be aware of the regression-to-mean effect. This phenomenon was already known from genetic research as the tendency to 'return to the mean': sons of very tall fathers are, on average, shorter than their fathers, and vice versa (Sir Francis Galton, 1877). A clear and extensive account of this phenomenon is Hauer & Persaud (1983).

1.5. Migration of Accidents

When an accident blackspot is treated, the reduction in accidents is often accompanied by an increase in numbers of accidents towards neighbouring sites. One explanation of this is a reduced awareness of the need for caution (Boyle et al., 1984), and risk compensation. The hypothesis of risk compensation has been criticised fiercely. Statistical explanations are the (reversed) regression-to-mean effect (McGuigan, 1985a), the possibility that conversions bring about a *degradation* in safety (instead of improvement), or positive correlation between true accident rates of adjacent sites, since traffic flow tends to be correlated at adjacent sites (Maher, 1987). Maher (1990, 1991) devised a bivariate negative binomial model to explain traffic accident migration, following Senn & Collie (1988), Bates and Neyman (1952). Persaud (1987) gives plausible explanations for migration of accident risk after remedial blackspot treatment.

1.6. Organisation of the Subjects

First of all, an overview of historical issues is presented (see Arbous and Kerrich, 1951), who popularised some standard distributions (and mixtures of distributions) for accident analysis. We discuss the concepts of accident proneness and accident liability, and the difference between the two. Some people believe in the existence of some sort of 'calamity Joe', but others argue that accidents occur at random to any experienced driver (section 2). Because accidents are aggregated for each *individual* road section, for example, and, in a second step, over *all* (comparable) road sections, estimation proceeds at two levels. With this kind of data, shrinkage estimators are more efficient; this follows from Stein's paradox (section 3). Some useful distributions for the analysis of accidents are given in section 4. Classical versus empirical Bayes techniques are the subject of section 5, and the pioneering work by Robbins with respect EB estimating techniques in section 6, and 7. Section 6 concerns estimation a compound Poisson distribution', in three distinct cases (G is unknown; G is restricted to be Gamma; and G is completely specified, resp.). Section 7 concerns prediction and estimation for mixtures of exponential distributions. The description of the Poisson-Gamma model and the estimation of the regression effect by Abbess, Jarrett & Wright (1981) triggered a lot of research, see section 8. The linear empirical Bayes estimator is presented in section 9. Estimating 'safety' is the subject of section 10, the correction for 'bias by selection' and the selection of truly dangerous site is the subject of section 11. The phenomenon of migration of accidents is discussed in section 12. Section 13 presents some special formulas for empirical Bayes estimation. References are given in section 14.

2. Accident Proneness, Compound Poisson & Mixtures

In the first half of the 20th century, a lot of research was being done in the field of human factors research. This research was concerned primarily with industrial accidents, but also with road accidents. Different hypotheses may give rise to the compound Poisson distribution, e.g., the *biased* distribution with special examples the *contagious* hypothesis and the *burnt fingers* hypothesis and the *unequal initial liabilities* hypothesis. An important review is by Arbous & Kerrich (Biometrics, 1951, ca. 90 pages). After this, Robbins (1956, 1964, 1977, 1980a,b,c) contributed influential papers about accident statistics. Hauer (1980a,b, 1986, 1992, 1995), contributed new directions, methodologies and applications in road safety research. Abbess, Jarrett & Wright (1981) suggested to use the negative binomial distribution to predict the future number of accidents for any blackspot from past and present accident frequencies. This paper was very influential and triggered a train of research, as did Jarrett, Abbess & Wright (1982), and Hauer (1980a,b). Hauer & Persaud (1983, 1984, 1987) and Persaud & Hauer (1983) are early contributions that should not be missed in the study of the methodology of road safety research.

2.1. History of Accident Statistics

In the study of accident proneness one has come to realise that "efficient and not cheap labour is the most economical" and that "full and promotive health is considered far more important than the mere absence of disease (...) in order to ensure maximum operating efficiency" (Arbous & Kerrich, 1951). This study is primarily concerned with industrial accidents but applies to road accidents as well.

As the authors note, in the period 1940-1945, South Africa suffered 23,300 war casualties and 54,987 casualties on the roads of the Union. It is realised that personal factors play a great part in accidents. Besides environmental circumstances, psycho-physiological circumstances (health, alcohol, fatigue, age and experience) affect the liability to accidents. The state of fatigue may be influenced by environmental circumstances (lack of rest pauses, food, illumination, etc.) and is considered to have great impact on both output of work and accident rate. Mental attitudes may well destroy the direct relationship between environmental factors and accidents. In the Hawthorne Plant of the Western Electric Company (Roethlisberger et al., 1949), a study was set up to examine the influence of various environmental factors (rest pauses, illumination, etc.). Production increased in both the test group and in the control group. In another experiment in the same plant, illumination was 70% decreased for the experimental group. As in the previous experiment, production increased in both the experimental and control group. The experimenters realised that *motivation* must be a factor. It was hypothesised that the new interest taken in the workers - they were kept fully informed and taken into the confidence of the investigators - caused this effect.

2.2. Accident Proneness

In most of the industrial work-groups studied, a minority were responsible for the majority of the accidents (Arbous & Kerrich, 1951). This does *not* imply that we clearly can discriminate between people who are accident-prone and who are not, as Kerrich (in Arbous & Kerrich, 1951) illustrates in Adelstein's (1951) data. In this study, the accident data of 104 shunters who joined in 1944 and shunted for three years. The average accident rate for 104 men in the 1st, 2nd, and 3rd years was: .557, .355, .317, resp. After removing the 10 men with the highest accident rate in the first year, the average accident rate for the remaining 94 men was: .393, .361, .329, resp.:

the accident rate increases somewhat after the removal of the men with highest accident rate. (If it is true that, on average, these men all had an equal accident rate, then the men with an early high accident rate are expected to have a relatively lower accident rate after the first year, and the shunters with low accident rate a *higher* accident rate.) Accident *proneness* is distinguished from accident *liability* (Farmer et al., 1939): "Accident *liability* includes all factors determining accident rate, accident *proneness* refers only to those that are personal". According to Vernon (1936), even accident-proneness is not a fixed quality of the individual, but is influenced by external changes of the environment as well as by internal changes of physical and mental health (see Arbous et al., 1951, p. 352). The statistical study of accident proneness started as early as 1919. The first study is by Greenwood and Woods (1919) who investigated the frequency with which accidents occurred in groups of 'munition' women engaged in various machine operations required in the manufacture of shells in a munition plant. Some of the women suffered no accidents at all, others suffered once or twice, and a few of them more frequently. This might be due to *simple chance*, or the workers may have started all equally, but an individual who suffered one accident by pure chance might, in consequence, have her probability of suffering further accidents increased or decreased. This might have made her more careful afterwards, which reduced her liability. Or it might have increased her nervousness, and thereby predisposed her to more accidents. Accidents distributed on this basis are called *biased* (see Arbous et al., 1951, p. 353). Or, we may suppose that all workers did not start equal, but some are more liable to suffer casualties than others, from the beginning. Accidents would then be distributed on the basis of *unequal liabilities*. The 'unequal initial liability' hypothesis has the closest fit to the data. Thus, three hypotheses are forwarded (Arbous et al., 1951):

1. *Hypothesis of Chance Distribution*

Accidents occur to individuals by pure chance. This means that individuals are equally liable to sustain accidents in the environment, and environmental circumstances are homogeneous for all individuals with regard to physical and psychological qualities that may influence liability. Then the distribution of accidents is Poisson. However, when the distribution of the group is Poisson, it need not be true that the population is homogeneous. For example, accidents in two distinct periods may be Poisson distributed in each period, but when accidents from the two periods are *combined*, the distribution may be a compound Poisson distribution.

2. *Hypothesis of Biassed Distribution: Burnt Fingers*

Different hypotheses may also give rise to the compound Poisson distribution, e.g., the biased distribution (Arbous et al., 1951, p. 358): the probability that an individual has an accident within a short interval of time depends on the number of accidents that he has had previously. One special example of the biased hypothesis is the '*contagious hypothesis*': the more accidents a person had, the more likely he is to have accidents (at time zero, none of the individuals have had an accident). Practical identical results follow from the so-called '*burnt fingers hypothesis*' ('the child who has burnt his fingers, fears the fire', p. 405), resulting in a *decrease* of accidents.

The hypothesis of increasing numbers of accidents due to increased liability after an accident cannot be the *right* one, because if it were true, accident rates of individuals should increase with time (the more accidents a person had, the more likely he is to have accidents, which is in contradiction with actual experience. But see Kerrich's treatment of the '*contagious hypothesis*' in Part II of Arbous et al. (1951), and below.

3. Hypothesis of Unequal Initial Liability - Accident Prone Percy

It is customary to choose the negative binomial distribution when the Poisson distribution is not appropriate. One of the hypotheses that lead to this type of distribution is that of unequal initial liability or proneness. Assume that n persons are working in a homogeneous environment, and that these persons are non-homogeneous with respect to their individual proneness to sustain accidents, i.e., they differ among themselves in this respect. Suppose that the total group can be divided into subgroups 1, 2, 3, ..., k , each of which is homogeneous within itself but differing from all the others. Although the liability in each subgroup is constant, individuals will not all have exactly the same number of accidents in any given period of time. By assuming homogeneity in each subgroup, the distribution of accidents will follow a Poisson distribution. Then it is possible to calculate the average number of accidents per individual (λ_i) in each subgroup, which represents the liability of individuals in each subgroup to sustain accidents. Thus,

Accidents of subgroup 1 yield a Poisson distribution with average λ_1 .
 Accidents of subgroup 2 yield a Poisson distribution with average λ_2 .
 : : : : : : : : : :
 Accidents of subgroup k yield a Poisson distribution with average λ_k .

The number of individuals per subgroup is assumed to vary. The form of the resulting distribution is a Gamma or a chi-squared distribution and the number of accidents by the total group, irrespective of subgroups, follows a negative binomial distribution (Arbous et al., p. 360), which comes down to a superimposition of a series of Poisson distributions, referring to the subgroups 1, 2, 3, ..., k , mentioned above. This is true as well for environments and personal liabilities or λ 's changing in time, provided the changes are the same for all individuals (Arbous et al., 1951, p. 362-363).

2.3. Compound Poisson: Mixture of Distributions

Let $\alpha = \lambda/T$ be the average accident liability, the average number of accidents per individual per unit time. If a population is not homogeneous, it may consist of a mixture of two or more homogeneous populations: a compound Poisson distribution. The distribution of the mixture is

$$f(x) = p_1(\lambda_1) e^{-\lambda_1} \lambda_1^x / x! + p_2(\lambda_2) e^{-\lambda_2} \lambda_2^x / x!, \quad p_1 + p_2 = 1, \quad (1)$$

and analogously for mixtures of $n > 2$ populations. Still more generally, suppose that the λ 's have a continuous distribution, so that the individuals with the same λ have a distribution with $f(x|\lambda) = e^{-\lambda} \lambda^x / x!$, while the distribution in the population is

$$f(x) = \int_0^\infty f(x|\lambda)G(\lambda)d\lambda, \text{ a mixture, denoted } f(x) = \int_0^\infty f(x|\lambda) g(\lambda), \quad (2)$$

in which the *conditional* Poisson probability $f(x|\lambda)$ is weighted by the *unconditional* probability $g(\lambda)$; $G(\lambda)$ is based on empirical frequencies.

Different hypotheses give rise to the compound Poisson distribution, e.g., the *unequal initial liabilities* hypothesis, the *biased* hypothesis with special examples the *contagious* hypothesis and the *burnt fingers* hypothesis.

2.3.1 Two Sub-Periods: Bivariate Compound Poisson Distribution

As for the 122 shunters, they had up to 25 years experience before coming under observation. They were observed for 11 years. This period was divided into two periods: from 0 to 6 years and from 6 to 11 years, respectively. Within each subinterval, a Poisson distribution can be fitted to the

observations, but not to the combined observations ($p = .15, .30, \text{ and } <.001$, respectively). The authors conclude that the population is non-homogeneous and they fitted a bivariate negative binomial distribution based on the compound Poisson hypothesis to the data, the fit seems quite satisfactory. The same has been done for the contagious hypothesis. For this data, the 'compound' hypothesis is more plausible than the 'contagious' hypothesis. The correlation coefficient between the periods is $r_{01} = .258$, which is significantly different from zero.

The fundamental question is to predict which people will have the most accidents in the future. This question may be researched in three stages, using the bivariate compound Poisson or the bivariate contagious distributions, respectively (Arbous et al., p. 413 - 425):

- (1) examination of what happened in the first period,
- (2) prediction of what will happen in the second period, and
- (3) evaluation of what happened in the second period.

Take two subintervals, δ_0 (the interval from 0 to t) and δ_1 (interval from t to T). Then, x_0 = number of accidents an individual has during the interval δ_0
 x_1 = number of accidents an individual has during the interval δ_1
 x = number of accidents an individual has during $\delta = \delta_0 + \delta_1$.

It is assumed that for each individual,

$$f(x_0 | \lambda) = e^{-\lambda\delta_0} \frac{(\lambda\delta_0)^{x_0}}{x_0!}, \quad \text{and} \quad f(x_1 | \lambda) = e^{-\lambda\delta_1} \frac{(\lambda\delta_1)^{x_1}}{x_1!}.$$

Thus, the underlying assumptions are not the same as before, where λ was a fixed annual accident rate unaffected by time. For the λ 's, a Gamma distribution is taken in the form

$$\rho(\lambda) = \frac{(\beta/\gamma)^\beta \lambda^{\beta-1} e^{-\beta\lambda/\gamma}}{\Gamma(\beta)},$$

and γ is a fixed *annual* average number of accidents per person. Number of years (periods) of observation is represented by the parameter α , intensity of accidents by β . A bivariate compound Poisson distribution is derived, where $\gamma\delta$ replaces λ ; see Arbous et al.:

$$f(x_0, x_1) = \frac{\beta^\beta \gamma^x}{(\beta + \gamma\delta)^{\beta+x}} \frac{\Gamma(\beta + x)}{\Gamma(\beta)} \frac{\delta_0^{x_0}}{x_0!} \frac{\delta_1^{x_1}}{x_1!}, \quad \text{for } f(x_0 | x_1), f(\lambda|x),$$

2.3.2. Estimating the (Bivariate) Negative Binomial Distribution

The negative binomial distribution and the bivariate negative binomial distribution together with the estimation of parameters and an account of previous research is clearly described in Arbous & Kerrich (1951). To estimate the parameters of the negative binomial distribution from a set of data, a computer program was written by Wyshak (1974). Jarrett, Abbess & Wright (1982; 1988) and Jarrett (1991) are clear expositions of (estimation of parameters in) the bivariate negative binomial model. Maher (1987) describes the truncated negative binomial distribution to model accident frequency data in order to arrive at a probabilistic interpretation of accident migration. Senn & Collie (1988) use the bivariate negative binomial model to separate the regression to the mean effect from the treatment effect. Maher (1990) fits a bivariate negative binomial model to explain traffic accident migration. Maher (1991) fits a modified bivariate negative binomial model for accident frequencies, in which the two true average accident rates may be correlated.

3. Shrinkage Estimators: Stein's Paradox

The best prediction about the future usually is obtained from the *average*, \bar{x} , of past results. However, Stein's paradox defines circumstances in which there are estimates better than simply extrapolating from the individual averages, namely,

$$z = \bar{x} + c(x - \bar{x}),$$

in which x is an individual average, and \bar{x} the grand average for all entities (individuals). The original example is in terms of batting averages (x) of 18 baseball players, rescaled into numbers between 0 and 1 (Efron & Morris, 1977). The constant c is a shrinking factor, $c < 1$. The paradoxical element in Stein's result is that it may contradict statistical theory. It has been proved that no estimation rule is better than the observed average. However, with three or more baseball players and the goal of predicting future batting averages for each of them, there is a better procedure than extrapolating from the individual averages (Efron & Morris, 1977). Stein's paradox states that the z -value, the James-Stein estimator of a player's batting ability, is a better estimate of the true batting ability than the individual batting averages x - thereby making the average an *inadmissible* estimator. For 16 out of the 18 players, the batting abilities are estimated more accurately than by calculating the individual averages. For these players, the individual average is inferior to the James-Stein estimator as a predictor of their batting ability. The James-Stein estimators, as a group, have smaller total squared error.

Let a player's *true* batting ability be denoted by θ ; an unobservable quantity, of which the performance of the batters is a good approximation. Now, it should hold that for most of the players, $|\theta - x| > |\theta - z|$. The total squared error of estimation is smaller for the James-Stein estimator (.022 for the James-Stein estimator and .077 for the observed averages). The explanation lies in the distribution of the random variable. Stein's (1955) theorem is concerned with the estimation of several unknown averages (the individual batting abilities) by a risk function (the sum of the expected values of the squared errors of estimation for all the individual averages). No relation between the averages need be assumed, and the averages are assumed to be independent. When the number of averages exceeds two, estimating each of them by its own average is an inadmissible procedure: there are rules with smaller total risk (Efron & Morris, 1977; James & Stein, 1961). The shrinkage factor is:

$$c = 1 - \frac{(k - 3) \sigma^2}{\sum (x - \bar{x})^2}, \text{ where}$$

- k is the number of unknown averages,
- σ^2 is the overall variance of x , and
- $\sum (x - \bar{x})^2$ is the sum of the individual variances (squared differences between individual averages w r t. the grand average)

The shrinking factor c becomes smaller as $\sum(x - \bar{x})^2$ becomes smaller. The procedure works best if the individual averages x are near the grand average \bar{x} . Shrinkage ('Linear Empirical Bayes', 'LEB') estimators are described in Robbins (1977, 1980a,c), Jarrett et al., (1988), Jarrett (1991, 1994), Persaud (1986, 1987).

4. Accident Distributions

The Poisson distribution has become a standard for the analysis of accident frequencies at a fixed location during a certain period of time (see e.g., Arbous & Kerrich, 1951; Robbins, 1956, 1964, 1977, 1980*a,b,c*; Johnson & Kotz, 1969; Gipps, 1980; Hauer, 1980*a,b*, 1981, 1982; Hauer & Persaud, 1983; 1984; 1987; Abbess, Jarrett & Wright, 1981; Jarrett, Abbess & Wright, 1982; 1988). Counts of (fatal) accidents are compared over periods and accident counts are aggregated over comparable locations using an empirical Bayes approach (see, e.g., Hauer, 1980*a,b*, 1981, 1986, 1992, 1995; Abbess, Jarrett & Wright, 1981; Hauer & Persaud, 1983; Persaud & Hauer, 1983). Counts of numbers of (fatal) accidents at a fixed location or at carefully matched locations are repeated over time, so there is variation in time and location. As for the locations, there are as many averages (and/or expectations) of numbers of accidents as there are locations. The averages/expectations themselves have a distribution with the *grand average* as a measure for central tendency and a dispersion factor. The distribution of the averages is supposed to be Gamma. If the distribution of accidents per location is Poisson, and if the distribution of the averages of the locations is Gamma, then the distribution of the accidents over locations is the negative binomial distribution.

4.1. Distributions: Poisson, Gamma, and Negative Binomial

There is always variation between locations (and between, for example, annual measurements at a fixed location) as a result of changes in traffic flow and other causes such as an adaptation of the traffic situation. As a result of this variation, frequencies of accidents at more locations and/or over more years will, in general, be not exactly Poisson distributed, but show 'overdispersion': the ratio (variance/average), the coefficient of variation, will not be exactly one, but larger than one.

To face this problem, the negative-binomial distribution is used to describe the distribution of accident frequencies (see e.g., Abbess, Jarrett, and Wright, 1981; Jarrett, Abbett, and Wright, 1982). The compound Poisson distribution is referred to as a 'Poisson distribution with overdispersion'. The negative binomial distribution arises from the combination of a Poisson distribution for X , the number of accidents at a certain location given $E(\lambda)$, the expected value of X , and a Gamma distribution to describe the variation of λ over locations (see Jarrett et al., 1982, Abbess et al., 1981; Robbins, 1950, 1951, 1956, 1964, 1977, 1980*a,b,c*; Hauer, 1981; Arbous and Kerrich, 1951; Bates & Neyman, 1952; among others).

Tables and properties of the negative binomial probability distribution are given by Williamson et al. (1963). A program for estimating the parameters of the truncated negative binomial distribution is Wyshak (1974).

5. Classical vs Empirical Bayes: Robbins (1956)

Both in pure Bayes and in empirical Bayes estimation, the parameter value λ itself is regarded as a realisation of the random variable Λ , with distribution function $G(\lambda)$, the *prior* distribution. In the pure Bayes approach, $G(\lambda)$ is not necessarily interpreted in terms of relative frequencies (it may reflect intuition or prejudice). In the EB approach, $G(\lambda)$ is given a frequency interpretation and the availability of previous data, suitable for estimation of $G(\lambda)$, is assumed. It was Robbins (1956) who coined the term 'empirical Bayes approach' (Maritz, 1970, p. 2). The EB estimation rule generally depends only on past and current accident frequencies.

Suppose there is an hazardous location, somewhere, and there is pressure from local authorities to make considerable adaptations. Of course, this is very costly and one has to make sure that major changes are called for. The problem is to estimate the *true* intensity, λ , of fatal accidents at that site, from the observed value, x_n , last year's intensity. Since accidents occur every year, there is a sequence of past observations (x_1, \dots, x_{n-1}) , the annual numbers of fatalities *before* last year. We suppose that there is a known probability density function $f(x|\lambda) = \Pr(X=x|\Lambda=\lambda)$ for the frequency of fatal accidents and a fixed distribution function for the true value λ : $G(\lambda) = \Pr(\Lambda \leq \lambda)$.

To estimate Λ_n from an observed X_n when the previous values $\Lambda_1 - \Lambda_{n-1}$ are known, the *empirical* distribution function of the random variable Λ , $G_{n-1}(\lambda)$ is used, which is based on the past $(n-1)$ observations. To ascertain the true risk of some site in comparison to other places or with respect to the past, the following sequence is generated (Robbins, 1956):

$$(\Lambda_1, X_1), (\Lambda_2, X_2), (\Lambda_3, X_3), \dots, (\Lambda_n, X_n)$$

The Λ_n are independent random variables with common distribution function G . The distribution of X_n depends only on Λ_n . For $\Lambda_n = \lambda$, $\Pr(X_n | \Lambda_n)$ is $f(x | \lambda)$. Let X_1, \dots, X_{n-1} represent the 'past' observed values of Λ , and let X_n be the 'present' value. Our task is to predict the unknown Λ_n from the observed value X_n . If the previous values $\Lambda_1, \dots, \Lambda_{n-1}$ are known, the empirical distribution of the random variable Λ can be formed, Λ_n can be estimated and the risk of the site can be evaluated. The present observation x is gauged with respect to the past information:

$$G_{n-1}(\lambda) = \frac{\text{number of terms } \Lambda_1, \dots, \Lambda_{n-1} \text{ which are } \leq \lambda}{n-1}, \quad (3)$$

the *frequency* or the number of years for which the number of fatalities was smaller than or equal to λ . For any fixed x the empirical frequency

$$f_n(x) = \frac{\text{number of terms } X_1 \dots X_n \text{ which equal } x}{n}$$

Last year's estimate is the *mean* of the posterior distribution of Λ given $X_n=x$:

$$\phi_G(X_n) = \frac{\int_0^\infty \lambda f(x|\lambda) dG_{n-1}(\lambda)}{\int_0^\infty f(x|\lambda) dG_{n-1}(\lambda)}, \quad (4)$$

from (3) by replacing the unknown prior distribution $G_n(\lambda)$ by the 'empirical' distribution $G_{n-1}(\lambda)$, and plugging in x , last year's estimate of λ .

5.1. Λ Stochastic

The variable X : 'number of deaths' takes discrete values $x = 0, 1, 2, \dots$ and has a probability distribution depending in a known way on the unknown Λ :

$$f(x | \lambda) = \Pr(X = x | \Lambda = \lambda); \quad (5)$$

Λ itself is a random variable with a prior cumulative distribution function

$$G(\lambda) = \Pr(\Lambda \leq \lambda). \quad (6)$$

The unconditional probability density function (p.d.f.) of X is

$$f_G(x) = \Pr(X = x) = \int_0^{\infty} f(x | \lambda) dG(\lambda) \quad (7)$$

(see Robbins, 1956, p. 157). The dependence of the marginal distribution of x , $f_G(x)$, on $G(\lambda)$ is indicated by the index G .

5.2. Mixtures: Mixed Poisson Distribution

For a given λ , the probability of x deaths in accidents at a roundabout in one year is

$$f(x | \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad (8)$$

(see 5) and in repetitions of the same process with different (but completely matched) roundabouts, the marginal distribution of fatalities at these roundabouts is the *posterior* or *mixed* Poisson distribution

$$f_G(x) = \int_0^{\infty} f(x | \lambda) dG(\lambda) = \int_0^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} dG(\lambda), \quad (9)$$

see (7). Here, $f_G(x)$ and $f(x | \lambda)$ are given distribution functions. $f_G(x)$ is called a 'mixture' of $f(x | \lambda)$ type distributions functions, with $G(\lambda)$ the *mixing distribution*; $f(x | \lambda)$ is called the *kernel* (Maritz & Lwin, 1989).

Suppose we want to estimate the unknown intensity of some roundabout, Λ_n , from the observed x_n . If the previous values $\Lambda_1, \dots, \Lambda_{n-1}$ are known, then the Bayes estimator of Λ corresponding to the prior distribution G of Λ in the sense of minimizing expected squared deviation is the random variable $\phi(x_n)$ (cf. Robbins, 1956):

$$\begin{aligned} \phi_G(x_n) &= \frac{\int_0^{\infty} \lambda f(x | \lambda) dG_{n-1}(\lambda)}{\int_0^{\infty} f(x | \lambda) dG_{n-1}(\lambda)} = \frac{\int_0^{\infty} \lambda^{x+1} e^{-\lambda} dG_{n-1}(\lambda)}{\int_0^{\infty} \lambda^x e^{-\lambda} dG_{n-1}(\lambda)} \\ &= (x+1) \frac{f_G(x+1)}{f_G(x)}, \end{aligned} \quad (10)$$

from (4). The combination of historical data and present information yields a posterior average which includes the information of both sources and, hence, is more reliable.

6. Estimation for Mixtures of the Poisson Distribution

Robbins (1977, 1980a,b,c) presents three distinct procedures to estimate $E(\lambda | x = a)$: the expected numbers of accidents on the basis of observed numbers of accidents, and provides prediction intervals for future numbers of accidents. The 1977 and 1980b papers are special because they concern the prediction of future accident numbers *smaller* than of *larger* than some critical value. These will be discussed in a separate section. The 1980a,c papers present three different cases for the distribution of the expected accident rates are considered: the distribution of the mean accident rates, $G(\lambda)$, may be Gamma, restricted to be Gamma, or completely unknown. If the prior distribution is Gamma, the estimator is a *Linear Empirical Bayes* estimator ('LEB'). However, if G is not Gamma, the T -estimator may not be consistent. A more general EB estimator, ' \tilde{T} ', is always consistent, but may be inefficient if G is Gamma. Robbins (1980c) proposes to combine both methods of estimation.

Let X be a random variable such that, given λ , a is Poisson distributed, $f(a|\lambda)$, with average λ , and λ has a prior distribution G . Our task is to estimate λ from the data. The three cases depend on whether the prior distribution G is

- completely unknown (Case a);
- restricted to be Gamma with unknown parameters (Case b);
- known to be Gamma with known parameters (Case c);

Case a , b , and c are discussed in sections 6.1, 6.2, 6.3, respectively.

6.1. Empirical Distribution Function G Unknown

In Robbins (1980a), the problem is to estimate the expected accident rate $E_a(\lambda | X = x)$, given the observed rate x . There are no assumptions concerning the distribution function G of λ . Let a be a fixed integer. The sample size is n . Assume that (λ_i, X_i, Y_i) , $i = 1, 2, \dots, n$, form a random sample from the distribution of (λ, X, Y) . The parameter values λ_i , the past values X_i , and the future observations Y_i are random variables; only the past values X_i are observable, the parameter values λ_i and the 'future' values Y_i are not. Given λ , X and Y are independent Poisson distributed with average λ . The problem is to estimate

1. $E_a = E(\lambda | X = a)$, the grand average, an unknown constant
2. $S = (\bar{\lambda} | X = a)$, the average of the λ 's for all X_i that equal a ;
3. $U = (\bar{Y} | X = a)$, the average Y -value for all X_i that equal a .

First estimate: $E_a = E_a(\lambda | X = a)$

Robbins (1980a,c) shows that with

$$\begin{aligned}
 f(x | \lambda) &= \Pr(X=x | \lambda) = e^{-\lambda} \lambda^x / x!, && \text{see (8), and} \\
 f_G(x) &= \Pr(X=x) = \int_0^{\infty} f(x | \lambda) dG(\lambda), && \text{so that} \\
 E(\lambda | X=x) &= \int_0^{\infty} \lambda f(x | \lambda) dG(\lambda) / f(x) = (x+1) \frac{f(x+1)}{f(x)}. && (11)
 \end{aligned}$$

from (4) and (10). Hence, $E(\lambda | X = a) = (a + 1) \frac{f(a+1)}{f(a)}$.

For $n \rightarrow \infty$, E_a is consistently estimated by $(a+1) \frac{N_{a+1}}{N_a}$, (12)

where 'N' denotes 'frequency'. Robbins (1980c) shows that T^\dagger ($\equiv h$) is the minimum mean squared error estimator of λ for $x = a$:

$$T = T(a, G) = E(\lambda | x = a) = \frac{\int \lambda f(a|\lambda) dG(\lambda)}{f(a)} = (a+1) \frac{N_{a+1}}{N_a}. \quad (13)$$

For large n , an approximately 95% confidence interval for T is given by

$$(a + 1) \cdot \frac{n(a + 1)}{n(a)} \pm \frac{1.96(a + 1)}{\sqrt{n(a)}} \sqrt{\frac{n(a + 1)}{n(a)} \left(1 + \frac{n(a + 1)}{n(a)}\right)}.$$

Second and Third Estimates: $S = (\bar{\lambda} | X = a)$, and $U = (\bar{Y} | X = a)$
To express S , U and T in symbols, Robbins (1980a, b) defines

$$V_i = \begin{cases} 1 & \text{if } X_i = a \\ 0 & \text{if } X_i \neq a \end{cases}, \text{ and } W_i = \begin{cases} X_i & \text{if } X_i = a + 1 \\ 0 & \text{if } X_i \neq a + 1 \end{cases}, \text{ and}$$

$N_a =$ the number of values of X_1, \dots, X_n that equal a ($a = 0, 1, \dots$). Hence,

$$S = (\bar{\lambda} | X = a) = \frac{\sum_{i=1}^n \lambda_i V_i}{\sum_{i=1}^n V_i} = \frac{\sum_{i=1}^n \lambda_i V_i}{N_a}, \quad U = (\bar{Y} | X = a) = \frac{\sum_{i=1}^n Y_i V_i}{\sum_{i=1}^n V_i} = \frac{\sum_{i=1}^n Y_i V_i}{N_a},$$

$$\text{and } T = \frac{\sum_{i=1}^n W_i}{\sum_{i=1}^n V_i} = (a + 1) \frac{N_{a+1}}{N_a}, \text{ and is used to estimate } E_a, S, \text{ and } U, \quad (14)$$

see (13). Robbins shows that, as $n \rightarrow \infty$:

$$\begin{aligned} \bullet \sqrt{n}(T - E_a) &\xrightarrow{D} N(0, \sigma_1^2), \\ \bullet \sqrt{n}(T - S) &\xrightarrow{D} N(0, \sigma_2^2), \\ \bullet \sqrt{n}(T - U) &\xrightarrow{D} N(0, \sigma_3^2), \end{aligned}$$

and $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$. An approximately 95% confidence interval for E_a is

$$T \pm 1.96 \sqrt{(a+1)^2 (N_{a+1}) (N_a + N_{a+1}) / N_a^3}, \text{ and analogously for } S \text{ and } U.$$

6.2. Restricted Empirical Bayes Method

In the previous paragraph, the statistic T was used to estimate E_a . In this paragraph, the 'restricted' Empirical Bayes approach (Robbins, 1980a) is presented, in which the distribution function G of λ is assumed to be Gamma. In this restricted approach, α and β are unknown parameters to be estimated from X_1, \dots, X_n ; the expected value is $E_a = \frac{a + \beta}{1 + \alpha}$ (see next section).

The test statistic is $\tilde{T} = \frac{a + \beta}{1 + \tilde{\alpha}}$ in which $\tilde{\alpha}$ and $\tilde{\beta}$ are estimates of α and β , based on X_1, \dots, X_n . Using the method of moments:

† Robbins' h is equivalent to Robbins' T (1980a). We stick to the notation ' T '.

$$\tilde{\alpha} = \frac{\bar{x}}{s^2 - \bar{x}}, \quad \tilde{\beta} = \frac{\bar{x}^2}{s^2 - \bar{x}}, \quad (15)$$

$$\text{where } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \text{so that} \quad (16)$$

$$\tilde{T} = \frac{\bar{x}^2 + (s^2 - \bar{x}) \alpha}{s^2} = \text{LEB}, \quad (17)$$

the Linear Empirical Bayes ('LEB') Estimator, see Jarrett, 1991. If G is Gamma distributed, then \tilde{T} will be more efficient than T , since $\sqrt{n}(\tilde{T} - E_a)$ is asymptotically normal with mean 0 and a smaller variance than T (Robbins, 1980a). However, if G is not Gamma, then \tilde{T} need not converge to E_a at all, in contrast to T , which always does.

6.3. Empirical Distribution G Completely Specified

1. The distribution function of the true underlying fatality rate Λ is:

$$\text{Gamma: } g(\lambda) = dG(\lambda) = \frac{\alpha^\beta}{\Gamma(\beta)} \lambda^{\beta-1} e^{-\alpha\lambda} \quad \alpha, \beta > 0$$

where • $\Gamma(\beta) = (\beta-1)! = (\beta-1)(\beta-2)(\beta-3)\dots(1)$,

$$\text{and} \quad \bullet E(\lambda^r) = \frac{\Gamma(r+\beta)}{\Gamma(\beta)\alpha^r} \quad (r \geq 0), \quad \text{so that} \quad E(\lambda) = \frac{\beta}{\alpha}, \quad \text{Var}(\lambda) = \frac{\beta}{\alpha^2} \quad (18)$$

$$\text{and} \quad \bullet \phi_G(\cdot) = \frac{x+\beta}{1+\alpha}, \quad \text{the posterior mean} \quad (19)$$

$$\text{and} \quad \bullet E[\phi_G(X) - \lambda]^2 = \frac{\beta}{\alpha(1+\alpha)}, \quad \text{the expected squared deviation}$$

2. The distribution of observed fatalities x given the true underlying rate λ is:

$$\text{Poisson: } f(x|\lambda) = e^{-\lambda} \lambda^x / x!, \quad x = 0, 1, 2, \dots$$

$$\text{with} \quad \bullet E(x|\lambda) = \text{Var}(x|\lambda) = \lambda$$

3. The mixed or posterior distribution, $p_G(a)$, a 'compound Poisson', is the

$$\text{Negative binomial: } f(a) = \frac{\Gamma(\beta+a)}{\Gamma(\beta)a!} \left(\frac{\alpha}{1+\alpha}\right)^\beta \left(\frac{1}{1+\alpha}\right)^a,$$

$$a = 0, 1, 2, \dots; \quad (\text{Robbins, 1980c, p. 6988}),$$

$$\text{with} \quad \bullet E(a) = E(\lambda), \quad \text{Var}(a) = E(\lambda) + \text{Var}(\lambda), \quad (20)$$

$$\text{therefore,} \quad \bullet E(a) = \frac{\beta}{\alpha}, \quad \text{Var}(a) = \frac{\beta(1+\alpha)}{\alpha^2}, \quad \text{from (18) and (20),} \quad (21)$$

$$\text{hence,} \quad \alpha = \frac{E(a)}{\text{Var}(a) - E(a)}, \quad \beta = \frac{E^2(a)}{\text{Var}(a) - E(a)}, \quad \text{from (20) and (21).} \quad (22)$$

6.4. Estimation for Compound Poisson: Various Statistics

Robbins (1980c) presents an estimator h for $E(\lambda | X = a)$ for the case where the empirical distribution function G is unknown. The h statistic is the same as Robbins' (1980a) T -estimator. Robbins (1980c) also presents a restricted empirical Bayes estimator T_n , for the case where G is restricted to be Gamma, and is the same as Robbins' (1980a) \tilde{T} -estimator. These statistics from Robbins (1980a) are discussed in a wider framework (see Table 1). Robbins (1980c) presents three new estimators, k , W_n , and Z_n . E_a denotes $E(\lambda|X=a)$, T and h are the consistent moment estimators of E_a ; \tilde{T} and T_n denote the restricted Bayes estimate when G is restricted to be Gamma; k is, like h , a moment estimator of $E(\lambda | X = a)$, but may be associated with a different prior distribution function, say G' . If, in fact, G is Gamma, then $\tilde{T} = T_n = T = h$, and \tilde{T} (or T_n) is more efficient in estimating $E(\lambda | X=a)$. But if G is not Gamma, \tilde{T} (or T_n) need not converge to $E(\lambda | X=a)$. Suppose that k is based on a prior other than G , and has the same form as

$$h = E(x) + \left(1 - \frac{E(x)}{\text{Var}(x)}\right) (a - E(x))$$

then h can be estimated by replacing $E(x)$ and $\text{Var}(x)$ by the consistent moment estimators from (13):

$$W_n = \frac{1}{n} \bar{x}_n + \left(1 - \frac{\bar{x}_n}{s_n^2}\right) (a - \bar{x}_n) \quad (\text{Robbins, 1980c}).$$

Using Z_n , these problems are overcome, because $Z_n \sim W_n$ when h (or T) = k ; $Z_n \sim T_n$ when h (or T) $\neq k$.

Table 1. Overview of statistics for estimating $E(\lambda | X = a)$ by Robbins (1980a, 1980c).

1980a:	E_a	T	-	\tilde{T}	-	-
1980c:	E_a	h	k	T_n	W_n	Z_n

Description of statistics

- $E_a = E(\lambda|X=a)$, expected value for λ given the observation $X=a$;
- $T \sim h = (a + 1) \frac{N_{a+1}}{N_a}$, consistent moment estimator of E_a ;
- $T_n \xrightarrow{p} h$; $T_n = \tilde{T}$ = 'LEB' = $\frac{\bar{x}^2 + (s^2 - \bar{x}) a}{s^2}$; $T = h = \frac{E(W)}{E(V)}$, see (13) - (17);
- $\tilde{T} = \frac{\bar{x}^2 + (s^2 - \bar{x}) a}{s^2}$, the 'restricted' or Linear Empirical Bayes estimator;
- $k \sim h$ ($\sim T$): the prior G of k may be different from that of h ;
- W_n is the LEB estimator based on the moment estimators (see (15) - (17));
- Z_n is a combination of T_n and W_n ; $Z_n \sim W_n$ when $h = k$; $Z_n \sim T_n$ when $h \neq k$, a 'super estimator': it combines the best of both estimators, T_n and W_n (Robbins, 1980c).

7. Prediction for Mixtures of the Exponential Distribution (Including the Compound Poisson)

An estimate of $E(\lambda | x > a)$, for some accident count a , will often be more informative than separate estimates of $\lambda_1, \dots, \lambda_n$. In Robbins (1977, 1980b), a special empirical Bayes approach is proposed for predicting the future number of accidents for people who endured less than x (or: more than x) accidents. Suppose that last year a randomly selected group of n people had x_1, \dots, x_n accidents. We are interested in the number of accidents y_1, \dots, y_n in the next year for these same n people. Let i be a positive integer, then the problem is to predict, from the values x_1, \dots, x_n alone, the value of

$$S_{i,n} = \{\text{sum of the } y_j\text{s for which the corresponding } x_j\text{s are } < i \},$$

the total number of accidents that will be experienced next year by those people who experienced less than i accidents last year. Suppose that we are interested in next year's accident rate for people who had zero accidents. If we take

$$\{\text{the sum of the } x_j\text{s that are } < i \},$$

as an estimator, this would underestimate $S_{i,n}$ by a considerable amount (Robbins, 1977). A good predictor of $S_{i,n}$ is

$$E_{i,n} = \{\text{sum of the } x_j\text{s that are } \leq i \} = \{\text{sum of the } x_j\text{s that are } = i \},$$

because $S_{1,n}$ is always zero (i.e., a sum of zeros). Therefore, a good predictor of $S_{1,n}$ is $E_{1,n} = \{\text{number of people who had exactly 1 accident last year}\}$. To prove this, Robbins used the compound Poisson distribution. The conditional probability distribution of x and y , given λ , is Poisson (see (8)):

$$f(x | \lambda) = e^{-\lambda} \lambda^x / x!, \text{ and } f(y | \lambda) = e^{-\lambda} \lambda^y / y!.$$

The average and variance of the conditional distribution are derived:
 $E(x | \lambda) = \lambda$, $E(x^2 | \lambda) = \lambda + \lambda^2$, $\text{Var}(x | \lambda) = \lambda$. (23)

The unconditional probability function of x and y is the compound Poisson:
 $f(x) = \int_0^\infty (e^{-\lambda} \lambda^x / x!) dG(\lambda)$ resp. $f(y) = \int_0^\infty (e^{-\lambda} \lambda^y / y!) dG(\lambda)$,

(see (9)), and for the unconditional distribution, the average and variance are
 $E_x = E\lambda$, $E(x^2) = E\lambda + E(\lambda^2)$, $\text{Var}(x) = E\lambda + \text{Var}(\lambda)$, (24)

and the same for y . Besides $S_{i,n}$ and $E_{i,n}$, we need the statistics $N_{i,n}$ and A_i :

$$\begin{aligned} S_{i,n} &= \text{sum of the } y_j\text{s for which the corresponding } x_j\text{s are } < i \\ E_{i,n} &= \text{sum of the } x_j\text{s that are } \leq i \\ N_{i,n} &= \text{number of the } x_j\text{s that are } < i \\ A_{i,n} &\text{ is a function of the } S_{i,n} \text{ and } E_{i,n}. \end{aligned}$$

Robbins' main theorem states that, for any fixed $i = 1, 2, \dots$, as $n \rightarrow \infty$,

$$\frac{(E_{i,n} - S_{i,n})\sqrt{n}}{(A_{i,n} \cdot N_{i,n})} \xrightarrow{D} N(0,1). \text{ The 95\% confidence interval for } S_{i,n}/N_{i,n} \text{ is}$$

$$\frac{E_{i,n}}{N_{i,n}} - \frac{1.96 A_{i,n}}{\sqrt{n}} < \frac{S_{i,n}}{N_{i,n}} < \frac{E_{i,n}}{N_{i,n}} + \frac{1.96 A_{i,n}}{\sqrt{n}}.$$

For large n , the endpoints of the confidence interval do not depend on the mixing distribution G . For finite n the exact probability does depend on G . $E_{i,n}$ can be used to estimate $G_i = E(\lambda | x < i) = E(y | x < i)$, since $E(y | \lambda) = \lambda$.

Thus,

$$- E(\lambda | x) = \frac{(x+1)f(x+1)}{f(x)}, \text{ and } E(\lambda | a) = \frac{(a+1)N(a+1)}{N(a)}, \text{ from (12)}$$

$$- E(\lambda | x \geq i) = E(y | x \geq i) = \frac{E\lambda - b_i}{1 - a_i}, \text{ with } a_i = \sum_{x=0}^{i-1} f(x), b_i = \sum_{x=0}^i x f(x) \quad (25)$$

$$\text{or one may use: } \frac{\text{sum of the } x\text{'s that are } > i}{\text{number of } x\text{'s that are } \geq i}. \quad (26)$$

$$- E(\lambda | x < i) = \frac{b_i}{a_i}, \text{ with } a_i \text{ and } b_i \text{ from (25)}$$

This establishes the prediction of future accidents. Analogous results follow for the negative binomial, the negative exponential, and for the normal distribution.

7.1. $E(\lambda | X > a)$: Mixtures of Exponential Distributions

Our task is to predict $E(\lambda | X > a)$. In the previous section, no assumptions were made about the underlying distribution function G of λ . In this section, the distribution of X is an unknown mixture of exponential variables with different means λ . Robbins (1980b) shows that from a random sample x_1, \dots, x_n of X values, $E(\lambda | X > a)$ can be estimated for any given $a > 0$. Therefore, it is possible to predict the average of all *future* observations taken on those X values that exceed a . Given λ , X and Y are independent exponential random variables, as in the previous section. Only the x_i are observed. Let $h'(a) = E(\lambda | X > a)$, and $v = \sum v_i$ is the number of terms x_1, \dots, x_n that exceed a . Let

$$- v_i = \begin{cases} 1 & \text{if } X_i > a \\ 0 & \text{if } X_i \leq a \end{cases}, w_i = v_i(x_i - a) = (x_i - a)^+ \text{ (size of positive difference);}$$

$$- E'_\alpha = E(\lambda | X > a) = \frac{E(w)}{E(v)}, \text{ which is estimated by } \frac{\sum w_i}{v} = (\text{i.e., } \frac{\sum \text{size}}{\text{count}})$$

The problem is to estimate

$$- E'_\alpha = \sum w_i / v \sim \text{mean size: } E(\lambda | X > a) \quad (27)$$

$$- S' = \sum v_i \lambda_i / v \sim (\bar{\lambda} | X > a): \text{ average } \lambda \text{ given } X > a, \text{ and} \quad (28)$$

$$- U = \sum v_i y_i / v \sim (\bar{Y} | X > a): \text{ average } Y \text{ given } X > a. \quad (29)$$

$$\text{Now, for } n \rightarrow \infty, \frac{\sum w_i / v - E'_\alpha}{\sqrt{\frac{1}{v} \left\{ \frac{\sum w_i^2}{v} - \left(\frac{\sum w_i}{v} \right)^2 \right\}}} \xrightarrow{D} N(0,1) \text{ (Robbins, 1980b).}$$

$\sum w_i / v$ is an estimator of $\sum w_i / v$, $\sum v_i \lambda_i / v$, $\sum v_i y_i / v$. Approximately 95% confidence intervals for $E'_\alpha = \frac{1}{v} \sum w_i$, $S' = \frac{1}{v} \sum v_i \lambda_i$, and $U = \frac{1}{v} \sum v_i y_i$ are:

$$(1) \quad E'_\alpha: \frac{\sum w_i}{v} \pm \frac{1.96}{\sqrt{v}} \cdot \sqrt{\frac{\sum w_i^2}{v} - \left(\frac{\sum w_i}{v} \right)^2}, \quad \text{from (27), and}$$

$$(2) \quad S': \frac{\sum v_i \lambda_i}{v} \pm \frac{1.96}{\sqrt{v}} \cdot \sqrt{\frac{1}{2v} \sum v_i \lambda_i^2}, \quad \text{from (28), and}$$

$$(3) \quad U: \frac{\sum v_i y_i}{v} \pm \frac{1.96}{\sqrt{v}} \cdot \sqrt{\frac{1}{v} \sum v_i y_i^2}, \quad \text{from (29).}$$

8. Regression to the Mean: Abbess, Jarrett & Wright

Gipps (1980) proposed a theoretical model which assumes that sites - in some area under study - have true average accident rates which follow a Gamma distribution. He derived a formula for predicting the magnitude of the regression effect for an individual site. Hauer (1980) addressed the problem of a group of sites (as Robbins did, see section 6), all having more than a given number of accidents, x_0 , in a 'before' period (and were treated together). Hauer did not make assumptions about the distribution of the *average* accident rates, he estimated the regression effect directly from accident rates. According to Abbess et al. (1981), Hauer's method is sensitive to random fluctuation. Abbess, Jarrett & Wright, following Gipps (1980) addressed the question of whether the regression to the mean effect is likely to be sufficiently large to be worried about. We follow Robbins' notation. Jarrett, Abbess, and Wright (1982) again used a different notation, as Hauer does, which does not add to the readability of these papers. Jarrett et al. changed notation, probably because Abbess et al. made a mistake in a formula, which was corrected by Jarrett et al. These papers and Wright et al. (1988), Jarrett et al. (1988), Jarrett (1991) are among the most frequently cited. As Abbess et al. (1981) note, Bayesian methods have been controversial among statisticians for a long time. They show that the Bayesian approach is well-suited to analysing accident blackspot data and providing information on which to judge the effectiveness of remedial treatment. In the following, the assumptions necessary to apply the Bayesian method are given.

Step 1. Likelihood

At each crossing, without remedial adjustments, accidents may or may not occur. We suppose that these accidents occur according to a Poisson process with a constant intensity of λ per year:

$$f(x | \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (x = 0, 1, 2, \dots)$$

Step 2. Prior

The *true* accident intensity λ varies over locations, the value for a certain location is unknown and is considered a random variable. Suppose that the *prior* distribution of λ can be described as a Gamma density $f_0(\lambda)$, with parameters n_0 and x_0 :

n_0 = number of years of observation
 x_0 = number of accidents in period n_0 .

Then $f_0(\lambda)$ is defined by
$$f_0(\lambda) = \frac{n_0(n_0\lambda)^{x_0-1} e^{-n_0\lambda}}{\Gamma(x_0)} \quad \lambda > 0 \quad (30)$$

with average x_0/n_0 , (β/α in Robbins' notation)
 variance $x_0/(n_0)^2$ (β/α^2 in Robbins' notation);

$f_0(\lambda)$ describes our knowledge of the accident frequency λ (Abbess et al., 1981; Jarrett et al., 1982). In Robbins' notation with $\beta \sim x_0$ and $\alpha \sim n_0$:

$$g(\lambda) = \frac{\alpha^\beta \lambda^{\beta-1} e^{-\alpha\lambda}}{\Gamma(\beta)} \quad \text{for } \lambda > 0. \quad (30)$$

Step 3. Observations

Data: n = number of periods of observation,

x = number of accidents in n periods, $x \sim$ Poisson with expectation λ .

When in a certain period x accidents are observed, the posterior distribution is Gamma with parameters n_1 and x_1 , where $n_1 = n_0 + n$ and $x_1 = x_0 + x$.

Step 4. Posterior

The posterior distribution is a Gamma distribution with parameters (n_1, x_1) :

$$f_1(\lambda) = f_1(\lambda | x) = \frac{n_1(n_1\lambda)^{x_1-1} e^{-n_1\lambda}}{\Gamma(x_1)} \quad \lambda > 0$$

$$\begin{aligned} n_1 &= n_0 + n, & n_1 &= n_0 + 1, \text{ for one period (one year, e.g.)} \\ x_1 &= x_0 + x & x &= \text{number of observed accidents.} \end{aligned}$$

For a period of one year ($n=1$), the posterior distribution is:

$$f_1(\lambda) = f(\lambda | x) = \frac{(n_0+1)^{x_0+x} \lambda^{x_0+x-1} e^{-(n_0+1)\lambda}}{\Gamma(x_0+x)} \quad \lambda > 0 \quad (31)$$

(Note that with $\alpha = n_0$ and $\beta = x_0$, Formula 30 is equal to 30' and Formula 31 is identical to Formula 31'):

$$g_1(\lambda) = \frac{(\alpha+1)^{\beta+x} \lambda^{\beta+x-1} e^{-(\alpha+1)\lambda}}{\Gamma(\beta+x)} \quad \lambda > 0. \quad (31')$$

Step 5. Prediction

If the prior distribution is Gamma, then the *predictive* distribution, the distribution of the number of accidents, a , in the *next* period, is a negative binomial distribution. According to Jarrett et al. (1982):

$$q(a) = \frac{\Gamma(x_0 + a)}{a! \Gamma(x_0)} \left(\frac{1}{1+n_0}\right)^a \left(\frac{n_0}{1+n_0}\right)^{x_0} \quad (a = 0, 1, 2, \dots), \text{ but} \quad (32)$$

$$q(a) = \frac{\Gamma(s_1 + a)}{a! \Gamma(s_1)} \left(\frac{1}{1+n_1}\right)^a \left(\frac{n_1}{1+n_1}\right)^{s_1} \quad (a = 0, 1, 2, \dots) \quad (32')$$

according to Abbess et al. (1981). [Abbess et al. write: $s_1 + a$ ($= s_0 + s + a$), see Formula (32'), which comes down to: $x_0 + x + a$ in our notation, and is not correct, it should be $s_0 + a$. In later publications, they refer to Jarrett et al. (1982), who use 'k' instead of s_0 and 'c' instead of n_0 . Substitution of k and c for x_0 and n_0 , resp., in the equation corresponding to (31) in Jarrett et al., yields the correct formula.] The variation in the number of accidents over locations can be described using $q(a)$, an *unconditional* distribution. If the accidents at different locations are independent random variables, and if $f_0(\lambda)$ is Gamma distributed, then $q(a)$ is the negative binomial distribution:

$$q(a) = \int_0^{\infty} f(a|\lambda) f_0(\lambda) d\lambda, \quad f_0(\lambda) \sim \text{Gamma}, \quad a \sim \text{Poisson}: f(a|\lambda) = \frac{e^{-\lambda} \lambda^a}{a!} \quad (33)$$

8.1. The Regression Effect

The posterior expectation is: $E(\lambda | x) = \frac{x_0 + x}{n_0 + 1}$ (or: $\frac{\beta + x}{\alpha + 1}$, in Robbins' notation). This is the best estimate of the true accident frequency λ , given the observed number of accidents, x . The posterior distribution reflects our uncertainty with respect to the value of λ . There is variation in the true accident rates, λ , of the locations with x accidents. This variation is around $E(\lambda | x)$, the expectation of the true accident rate λ after having observed x accidents.

The difference between $E(\lambda | x)$ and x is the *regression effect*: $E(\lambda | x) - x$, which may be a positive or a negative effect. (Jarrett et al., 1982, 1988; Jarrett, 1991, 1994), see *Figure 1*. To determine the size of the regression effect, x_0 and n_0 must be known. The regression-effect is positive (i.e., more accidents than expected) if $x < x_0/n_0$, negative if $x > x_0/n_0$ (less accidents than expected). For $x = x_0/n_0$, the regression line is the 'line of expectations' and is a linear function of x , with slope $1/(n_0 + 1)$. The parameter n_0 necessarily is positive, therefore, the slope is always smaller than one (see Jarrett et al., 1982). Estimation of x_0 and n_0 from known accident data, is an *empirical* Bayes approach. Determining x_0 and n_0 , such that the prior distribution reflects a *subjective* opinion about the accident frequency, is the *subjective* Bayes approach and is not pursued here.

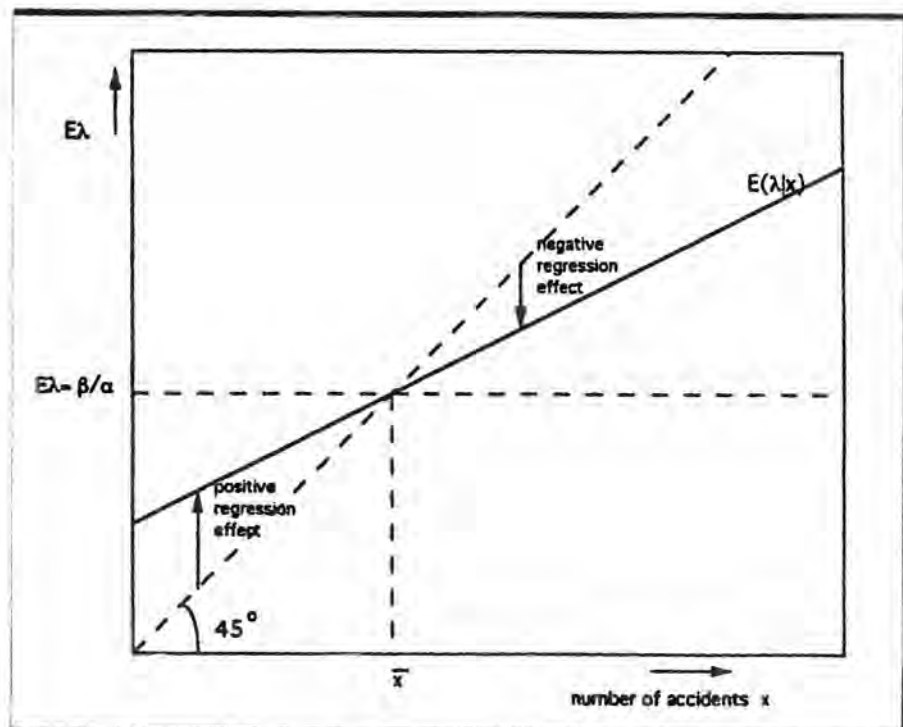


Figure 1. The Regression Effect.

9. Linear Empirical Bayes (LEB) Estimation

The observed accident frequency need not be the best estimator of the underlying mean. The so-called 'naive' method takes

$$\hat{\lambda} = x \quad \text{or} \quad \hat{\lambda}_{\text{naive}} = x.$$

Quite a different method is the 'Linear Empirical Bayes' method (first described by Robbins, 1977, 1980a,b,c), which yields the 'LEB'-estimator (Jarrett, 1991, 1994) for λ . According to Jarrett (1991), the LEB estimate can be better understood as a regression equation: the LEB formula gives an estimate of $E(\lambda | x)$, the regression of λ on x . Given the accident frequencies in one period for a group of locations, the estimator of the average, λ , at a location with x accidents is

$$\hat{\lambda}_{\text{LEB}} = \text{est. } E(\lambda | x) = \frac{\bar{x}^2}{s^2} + \left[\frac{s^2 - \bar{x}}{s^2} \right] x.$$

Here, \bar{x} is the grand average of the group and s is the standard deviation of the accident frequencies over all sites. [If all λ 's are equal, then $s^2 = \bar{x}$; also, if $\bar{x} = s^2$, then $\bar{x} \cdot \frac{\bar{x}}{s^2} + x \left(1 - \frac{\bar{x}}{s^2}\right) = \bar{x}$.]

The LEB-estimator is suited if the accident frequency x is Poisson distributed with average λ , for each separate location, independently of all other locations, and if the values of λ for the different locations can be considered as independent draws from a population with grand average \bar{x} and variance $s^2 - \bar{x}$. If further assumptions are made about the distribution of λ , better empirical Bayes estimates may be obtained, see Abbes et al. (1981), Jarrett et al., 1982. In fact, the LEB estimator is equal to Robbins' (1956, 1980a,b,c) T -estimator (see sections 7-9).

The LEB-estimate pulls or 'shrinks' x towards the estimated population mean. This results in a smaller total quadratic error, $\Sigma(\text{error})^2$, than for the naive estimators. Efron & Morris (1977) show that the James-Stein estimator $z = \bar{x} + c(x - \bar{x})$ gives better estimates of the true batting ability for baseball players than the individual batting averages (c is the shrinking factor, see section 3). The James-Stein estimator does substantially better than the averages if the true means lie near each other and can be used without any prior information. In historical context, the James-Stein estimator can be regarded as an 'empirical Bayes rule' (Efron & Morris, 1977, p.127). The practical value of the LEB estimator is that a better estimate can be achieved of the underlying mean, so that the treatment effect can be better determined. The LEB-formula yields an estimate of the regression line (see section 1.3 and section 8).

9.1. Two-Period Regression: 'Before' and 'After'

Another method to estimate the true accident intensity is the 'two period regression method' (Jarrett, 1991), in which a regression model is fitted to accident frequencies at untreated locations in two separate periods. These periods should be equal to the before and after period of the treated location. The 'before' frequencies are denoted x , and the 'after' frequencies y and a line, $\hat{y} = a + bx$, is fitted to the data. The results can be used to predict the accident frequencies, \hat{y} , that would have been expected had the location not been treated. Thus, the effect of the treatment, $\hat{y} - y$, can be estimated from $(a + bx) - y$, a relative reduction of

$$\text{reduction in \%} = \frac{\hat{y} - y}{y} \times 100\% = \frac{(a + bx) - y}{a + bx} \times 100\%.$$

From the principle of *regression to the mean* one expects this reduction to be smaller than the

$$\text{'naive' reduction} = \frac{(x - y)}{x} \times 100\%.$$

For a group of treated sites, the total reduction in numbers of accidents at the treated sites can be compared with the total predicted number of accidents had the treatment not been carried out:

$$\text{reduction in \%} = \frac{\sum(a + bx) - \sum y}{\sum(a + bx)} \times 100\%,$$

where the summation is over the treated sites.

9.2. Computing the Regression Effect

Abbess et al. (1981), show the effect of *not* including information from the past. They present accident frequencies from Hertfordshire (England) and calculate the size of the regression to the mean using a prior which is based on one year, two years, etc. The *percentage change* R is given by

$$R = \left(\left(\frac{x_0 + x}{n_0 + n} \right) \frac{n}{x} - 1 \right) \times 100\%, \quad (34)$$

where x_0 and n_0 represent information from the past (x_0 = number of accidents in the before period, n_0 = length of before period); x and n represent actual information: x = number of accidents during the present period, n = length of present period).

When the prior is based on only one year ($n_0 = 1$), the regression effect is largest, and diminishes when the prior distribution is based on increasing numbers of years of observation. The size of the regression effect in this example may be larger than 20 percent (after 1 year of observation), and may diminish to about 11% (after 2 years) and to about 8% in three years. The average number of accidents is 762, the average number of sites included per year is 270 (Abbess et al., 1981).

10. Estimating Safety

In the context of managing the 'unsafety' of traffic systems, two fundamental questions arise (Hauer & Persaud, 1987): first, how to identify systems that are unsafe, and secondly, how to estimate the safety effect of warning devices, such as crossbucks, traffic lights, etc.

Safety is a property of a specific entity, a crossing or a driver, etc. 'Safety' is defined as the number of accidents and their adverse consequences expected to occur, in the long run, with relevant conditions unchanged. The 'unsafety' of a crossing, λ , is unknown, but can be estimated by $\hat{\lambda}$. For a group of similar crossings, the average of their λ 's will be denoted by $E(\lambda)$, the 'grand average' and the variance of their λ 's is denoted $\text{Var}(\lambda)$.

The observed number of accidents, x , at a fixed time of the day and at a fixed location is supposed to be Poisson distributed with average λ . In each period, each crossing is characterized by its own λ . Each crossing has unique characteristics. Apart from these, it is possible to distinguish crossings with comparable characteristics, these crossings form a *group* of crossings.

From the number of observed accidents x , the average λ is computed. The average number of accidents is a measure of the (un)safety of that particular traffic system. Crossings or rail-highway grade crossings with the same or comparable characteristics such as traffic flow, number of lanes, type of crossing etc., may be treated as a group, with

- $n(x)$ is the number of crossings in the group, with x accidents, $x = 0, 1, 2, \dots$
- $E(\lambda)$ is the average value of the group,
- $\text{Var}(\lambda)$ is the variance of the group.

Since we are not concerned with exactly one Poisson distribution, but with a *mixture* of Poisson distributions, this mixture has a larger dispersion than the Poisson distribution, and it holds that (see Robbins, 1980c; Hauer, 1986; Hauer & Persaud, 1983; Persaud & Hauer, 1984)

$$\begin{aligned} E(x) &= E(\lambda) && \text{(from (24))} \\ \text{Var}(x) &= E(\lambda) + \text{Var}(\lambda) . \end{aligned}$$

Furthermore,

$$\begin{aligned} \bar{x} &= \Sigma [x \cdot n(x)] / \Sigma n(x) \\ s^2 &= \Sigma [(x - \bar{x})^2 \cdot n(x)] / \Sigma n(x) \\ \hat{\sigma}(\lambda) &= (\text{Var}(\lambda))^{1/2} = (s^2 - \bar{x})^{1/2}, \text{ since } \text{Var}(x) = E(\lambda) + \text{Var}(\lambda). \quad \text{(from (24))} \end{aligned}$$

Example 1 (taken from Hauer, 1992)

The distribution of the number of accidents in a homogeneous group of rail-highway grade crossings, defined by {urban area, 1 track, 1 to 2 trains per day, 0 to 1000 vehicles per day, crossbucks}, is characterised by $\hat{\sigma}(\lambda) \cong 2\lambda$, hence, the estimate of the standard deviation of the average is twice the expected value. Thus, even in a homogeneous group of such crossings we cannot assume that all λ 's are equal. It might be, that with larger groups, the variance of the average, $\text{Var}(\lambda)$, will tend to zero, so that the group may be characterised by one value, the group average. Using the results of an analysis of variance, this turned out to be not the case for this group. From this, Hauer concludes that the λ 's have a distribution with positive variance.

10.1. Poisson-Gamma model

If the distribution of the averages λ in a group of systems (crossings, drivers, etc.) can be described by a Gamma distribution, then the frequencies

$$a = n(x)$$

follow a negative binomial distribution, see Example 1. The reverse is also true: if the counts a follow a negative binomial distribution, then the averages are Gamma distributed (Jarrett et al., 1982). The fit of the negative binomial distribution to Hauer's data is close. Another example is the Hertfordshire data (Abbess, Jarrett, and Wright, 1981, p. 538-541). The authors not only describe the goodness-of-fit of the model, but they also give an account of the phenomenon of regression to the mean. Using the negative binomial distribution, tables for the test statistic, and data over more years for one location, these authors illustrate the phenomenon of 'regression to the mean'. This model yields a *criterion* for the *increase* of the number of accidents per year.

10.2. Parameter Estimation

Hauer & Persaud (1984) show that, if for each crossing, $x \sim \text{Poisson}$ and for each group of crossings $\lambda \sim \text{Gamma}$, then the expected value for a crossing can be best estimated using

$$\begin{aligned}\varepsilon &= x + [E(\lambda) / (E(\lambda) + \text{Var}(\lambda))] [E(\lambda) - x] \\ &= \alpha E(\lambda) + (1 - \alpha)x, \text{ in which} \\ \alpha &= (1 + \text{Var}(\lambda) / E(\lambda))^{-1} \quad \text{(from (18))}\end{aligned}$$

(from 11.1 and 11.2). Here, ε is a mixture of observed (x) and expected ($E(\lambda)$); α is a weighting factor, $0 \leq \alpha \leq 1$.

If $\text{Var}(\lambda) \gg E(\lambda)$, the group is heterogeneous (different averages);
small weight for data: α is small and $\varepsilon \cong x$

If $\text{Var}(\lambda) \sim E(\lambda)$, then homogeneous in the averages, $(1-\alpha)$ is small, $\varepsilon \cong E(\lambda)$.

A different estimation procedure is by using multivariate analysis: Hauer (1986; 1992), using GLIM (Aitkin et al., 1989), the negative binomial distribution.

11. Bias by Selection: Hauer, Hauer & Persaud

Hauer (1980a) describes the overestimation of the effectiveness of countermeasures when systems (intersections, drivers, vehicles, etc.) are selected on the basis of their lack of safety. This was called the *regression effect* (Jarrett et al., 1982; McGuigan, 1985a): systems that have, in some year, accident rates above average tend to have lower accident rates in the next year. Hauer presents a method to estimate this bias by taking the accident history into account: p_{ij} is the probability that system i has j accidents. Out of n systems (e.g., intersections) a number have to be selected for countermeasure implementation because they appear to be hazardous. Those systems are chosen that during some period of time t had x or more accidents. With

$B(x)$: expected # accidents during t before countermeasure (B : 'Before'),
 $A(x)$: expected # accidents during T after countermeasure (A : 'After'),

and when the countermeasure has no effect, the expected 'apparent' reduction of accidents per unit of time (with constant expected number of accidents) is:

Bias: difference $D(x) = B(x)/t - A(x)/T$

the expected *apparent* reduction in numbers of accidents per unit of time ($D(x) = B(x) - A(x)$ for equal t and T), assuming that the true number is equal in the before and after period. This is the phenomenon of 'regression to the mean'.

11.1. 'Before' and 'After' Period are Equal

If the 'selection bias' is not eliminated from the results, the effectiveness of the countermeasure may be overestimated. The correction of the bias follows from the comparison of an unbiased estimate of $A(x)$ and the actual number $A(x)$. Define

- the expected number of accidents at system i is λ_i ($i=1, \dots, n$)
- the probability that system i has j accidents is p_{ij}
- the probability that system i is chosen = $\sum_{j=k}^{\infty} p_{ij}$ (34)

- the contribution of system i to $B(k)$ is $\sum_{j=k}^{\infty} j p_{ij}$ (35)

Summation over all systems gives $B(k) = \sum_{i=1}^n \sum_{j=k}^{\infty} j p_{ij}$ (36)
 and p_{ij} is Poisson distributed. Thus,

$$B(k) = \sum_{i=1}^n \lambda_i \sum_{j=k-1}^{\infty} p_{ij} \quad (\text{Hauer, 1980}) \quad (37)$$

When the countermeasure has been implemented, the expected number of accidents at system i is equal to λ_i , hence, for all systems

$$A(k) = \sum_{i=1}^n \lambda_i \sum_{j=k}^{\infty} p_{ij} \quad \text{hence} \quad A(k) = B(k+1). \quad (38)$$

(Hauer, 1980; Hauer & Persaud, 1983). \hat{B} is an estimate of the accident history 'before', just as \hat{A} stands for 'after'. $B(k)$ includes all systems that exceed some critical threshold: the number of accidents is larger than or equal to k .

11.1.1. Size of Selection Bias

Hauer presents an example of a large sample of road sections from which six were selected to be resurfaced. They were selected because their three years accident totals prior to treatment were 15 or more: 15, 15, 16, 18, 18, 23, and 27. Thus $\hat{B}(15) = 132$ accidents. Without remedial treatment, one should expect for the next three years $\hat{A}(15) = B(16) = 0 + 0 + 16 + 18 + 18 + 23 + 27 = 102$ accidents, hence the bias $D(k) = 132 - 102 = 30$, a sizeable effect. The correction for the selection bias is

$$\% \text{ correction} = \frac{B(x) - A(x)}{\hat{B}(x)} \times 100\%$$

the percentage regression-to-the-mean. In general, with increasing observation periods, the bias decreases: $\lim_{k \rightarrow \infty} D(k) = 0$ (Hauer, 1980a; Wright et al., 1988).

11.1.2. Selection Bias is One-Sided

While the estimator, which eliminates the selection bias, is unbiased, it may not be precise (Hauer, Byer & Joksch, 1983). The bias of the original, biased, estimator is always in one direction. Use of the biased estimator may result in accepting ineffective treatment as effective. The unbiased estimator has a large variance: $E(\hat{\lambda}) = \lambda$ and $E(\hat{\lambda}_b) = \lambda + \text{bias}$ (see Hauer et al., p. 324). The biased estimator $\hat{\lambda}_b$ may have a lower mean squared error than the unbiased estimator $\hat{\lambda}$. However, for sufficiently large n the unbiased estimate will have a smaller mean squared error than the biased estimate. Hauer et al. give estimates of the variances of the biased and the unbiased estimators and other statistics. In the above example, the unbiased estimator would lead to a two-sided 90% confidence interval of [66 - 138], the biased estimator to [114 - 150]. Thus, with an 'after treatment' accident record of 110, the test using the biased estimator would have led to the conclusion of a significant improvement, whereas the test using the unbiased estimator would not. On the other hand, with an 'after treatment' accident record of 145, the test using the biased estimator would have led to the conclusion of no significant change, whereas the test using the unbiased estimator would have led to a significant worsening of the situation. Because the bias is always in the same direction, use of the *biased* estimator will, in the long run, result in *ineffective treatment being accepted as effective* (Hauer, Byer & Joksch, 1983).

11.2. The Shaman

Hauer and Persaud (1983)'s 'Common Bias in Before and-After Accident Comparisons and Its Elimination', shows that the selection bias is comparable to the 'all-pervasive and empirically substantiated phenomenon of regression to the mean' (Sir Francis Galton, 1877): tall fathers tend to have, on average, shorter sons and vice versa. The parallel in traffic safety is easy. Intersections, drivers, etc., are often selected for improvement on behalf of a bad safety record. Due to the principle of regression-to-the-mean, one should expect these systems to have less accidents in a subsequent period even if nothing is done to them. Yet in a simple before-and-after study, such an observed reduction is normally interpreted as indication that a countermeasure has been effective. To quote Hauer and Persaud (p. 164):

"Imagine now what would have happened had the Ministry hired a shaman to pray for accident reduction on road sections that had in the first year seven or more accidents. The apparent effectiveness of this 'treatment' is (...) about 30 percent."

11.2.1. *Expected Number of Accidents*

When the length of the before period increases, the relative size of the regression effect diminishes (cf. Abbess et al., 1981; Jarrett et al., 1982), because relatively large (or small) effects tend to average out in the long run. Hauer & Persaud raise the question of how to obtain unbiased estimates in uncontrolled before-and-after designs. Let γ denote the number of accidents occurring on a system during some period before treatment and let δ denote the number of accidents expected to occur on the same system during an after period of equal length had treatment not been applied. The problem is to estimate γ . Let $B(k)$ be the expected number of accidents on the chosen systems in period 1 ('Before') and $A(k)$ be the expected number of accidents on the chosen systems in period 2 ('After'), Hauer and Persaud derive three more or less equivalent estimation rules (from 12.1 - 12.5), namely:

$A(k) = B(k+1)$: the number of accidents expected to occur in the after period on systems that during the before period had k accidents is the number of accidents occurring on systems that had $(k+1)$ accidents in the before period. This rule leads directly to the third estimation rule.

$A(k) - A(k+1) = B(k+1) - B(k+2)$: the number of accidents expected to occur in the after period on a system that during the before period had k accidents is the number of accidents occurring on systems that had $(k+1)$ accidents divided by the number of systems on which k accidents occurred in the before period. This rule leads to $\gamma(k) = \delta(k+1)$.

$\gamma(k) = A(k) - A(k+1)$ and $\delta(k) = B(k+1) - B(k+2)$ the number of accidents expected to occur in the after period on systems that during the before period had k or more accidents is the number of accidents occurring on systems that had $(k+1)$ accidents or more in the before period.

The $\gamma(k)$ may be counts, relative frequencies, or cumulative frequencies. Accident data from different countries, relating to road sections, intersections, traffic circles, driver violations, driver accidents, and golf players are used to illustrate (the estimation of) the size of the bias, the technique for its elimination, and to examine the success of the debiasing procedure. Explanations are given for disproportional decreases, e.g., drivers with one or more violations or accidents in a first period are observed to have, on average, 60 - 80 percent less violations or accidents in the second period, in spite of the increase in the number of violations and accidents from the first to the second period. This change is much larger for drivers than that for elements of the road system. Furthermore, the smaller the average towards which observations regress, the larger the change (p. 167).

11.3. *Empirical Bayes vs Nonparametric*

Persaud and Hauer (1984) compare two methods of debiasing, a Bayesian and a nonparametric approach. The Bayesian method was found to yield better estimates. The nonparametric method overestimates the observed number of accidents or violations per driver. For systems with zero or one accidents/violations, the nonparametric method performs at least as well as the Bayesian method, which is an important result, because systems with $k = 0$ or $k = 1$ accidents are often of interest. By smoothing the frequencies, the Bayesian method yields the more reliable estimates for smaller groups of systems. Use of control groups yields better estimates of the number of after period accidents without treatment.

11.3.1. *Nonparametric Method*

The nonparametric approach is simple to apply and *only* assumes a Poisson distribution for the accidents (see Robbins, 1956). To estimate the number of accidents, a_k , expected to occur in an equivalent after period on a system that had k accidents in the before period, we need

N_k = number of systems with k accidents, and
 N_{k+1} = number of systems with $(k+1)$ accidents. Now,

$$a_k = (k + 1) \frac{N_{k+1}}{N_k}. \quad (\text{from (11)})$$

11.3.2. *Empirical Bayes Method*

The empirical Bayes method is just as simple as the nonparametric one. Its assumptions are (1) a Poisson distribution for the accidents, and (2) a Gamma distribution for the distribution of the *averages* in the population of systems. With these two assumptions, the number of systems with k accidents must obey the negative binomial distribution. The expected number of accidents, a'_k , in the after period on a system that had k accidents in the before period is

$$a'_k = (k + 1) \frac{N'_k}{N_k}$$

N'_k is the number of systems expected by the negative binomial distribution to have k accidents (see Robbins (1956) and sections 6 and 7). The before period accident data are used to estimate of the sample average, λ , and sample variance, s^2 for the population of systems. From these, parameter estimates for β and α of the Gamma distribution are obtained (cf. Robbins, 1980a):

$$\beta = \frac{\lambda^2}{s^2 - \lambda}, \text{ and } \alpha = \frac{\lambda}{s^2 - \lambda} \quad (\lambda < s^2) \quad (\text{see (18)})$$

(also, see Robbins, 1956; Jarrett et al., 1982; and Abbess et al.; 1981). If the negative binomial distribution fits the observed frequencies perfectly, the two methods, empirical Bayes and nonparametric, give identical estimates.

11.3.3. *Comparison of Both Methods*

From their empirical research, Hauer & Persaud (1983) conclude that for both driver and road systems, the Bayesian method is likely to give somewhat better estimates. For the drivers, the nonparametric method consistently overestimates the number of accidents or violations per driver. For systems with zero or one accident, the nonparametric method is slightly better. For any value of k larger than 1, the Bayesian method is better.

11.4. *The Sieve*

To identify hazardous locations or blackspots, a two-stage process is used by Hauer and Persaud (1984). The first stage acts as a sieve. A good sieve is one that allows through all sites that do not require remedial activity and retains all sites that do require detailed study. An inefficient sieve is one that retains a large number of sites that do not need close scrutiny and allows blackspots through its holes (Hauer et al., 1984). The quality and performance of a common sieve is evaluated in terms of a decision theoretic framework. What one hopes to find is the average number of accidents in the long run,

for each site. The difficulty is that accident histories are subject to substantial random fluctuation. The quality of the sieve can be evaluated by the

1. Number of ('#') sites selected (measure of effort, 'S')
2. # truly deviant sites among those selected (correct positives, 'CP')
3. # sites captured by the sieve but not deviant (false positives, 'FP')
4. # truly deviant sites not captured by the sieve (false negatives, 'FN')

The number of sites selected for inspection is CP + FP. These are the sites for which the number of accidents, x , exceeds a critical value, x^* : $x > x^*$. Truly deviant are sites for which $\lambda > \lambda^*$, the CP's and the FN's (not selected). Now, $n(x)$ is the frequency of x ; $F(\lambda | x)$ (see the Appendix) is the probability that the expected value of the number of accidents at a site is less than or equal to λ when the number of accidents actually recorded is x . $F(x | \lambda)$ is known for every case, $F(\lambda | x)$ is not. Knowledge of $n(x)$ and $F(\lambda | x)$ will enable to gauge the performance of the screening process, the sieve. How to obtain $F(\lambda | x)$, is discussed below.

11.4.1. Sieve Efficiency

According to Bayes' theorem, $f(\lambda | x) \propto f(x | \lambda) dG(\lambda)$. Integration of $f(\lambda | x)$ yields $F(\lambda | x)$. The coefficient of proportionality is selected such that

$\int_0^{\infty} F(\lambda | x) d\lambda = 1$; $f(x | \lambda)$ is assumed to be Poisson distributed and the numbers $G(\lambda)$ are ascertained from $n(x)$, the (empirical) frequencies of sites with x accidents. This is because the number of sites with x accidents, $n(x)$, must be a reflection of the distribution of λ among all sites (Hauer & Persaud, 1984). Therefore, the probability that a random selected site has x accidents is:

$$E\left\{ \frac{n(x)}{\sum n(x)} \right\} = \int_0^{\infty} f(x | \lambda) dG(\lambda). \quad (\text{see (2)})$$

(see section 3.2). When $G(\lambda)$ is a Gamma distribution, and x is Poisson distributed, the probability that a random selected site has x accidents obeys the negative binomial distribution. The parameters of $G(\lambda)$ can be estimated from the sample average and variance (12) and (13), the LEB-estimators:

- 1) Determine sample average and variance (summation is over all values of x)

$$\bar{x} = \Sigma [x n(x)] / \Sigma n(x), \quad s^2 = [\Sigma (x - \bar{x})^2 n(x)] / \Sigma n(x)$$

- 2) Estimate parameters α and β and determine $G(\lambda)$:

$$\hat{\alpha} = \bar{x} / (s^2 - \bar{x}), \quad \hat{\beta} = \bar{x}^2 / (s^2 - \bar{x}), \quad G(\lambda) = \alpha^{\beta} \lambda^{\beta-1} e^{-\alpha\lambda} / \Gamma(\beta), \quad \text{when } \lambda \geq 0.$$

It follows that $f(x | \lambda) \propto \lambda^{x+\beta-1} e^{-\lambda(1+\alpha)}$, a Gamma density. The posterior density $f(\lambda^* | x)$ is a Gamma density with α and β known. $F(\lambda^* | x)$ may be found by using numerical integration of

$$\frac{\int_0^{\lambda^*} \lambda^{x+\beta-1} e^{-\lambda(1+\alpha)} d\lambda}{\int_0^{\infty} \lambda^{x+\beta-1} e^{-\lambda(1+\alpha)} d\lambda}. \quad (\text{see (8)})$$

$F(\lambda^* | x)$ is the posterior probability that a site is truly deviant given the observed number of accidents. The probability distribution function depicted against λ , shows, for each value of λ , the number of systems that have more than x accidents per year. From this, the number of correct (false) negatives and positives can be ascertained. The number of sites to be expected, the numbers of correct positives, false positives and false negatives serve as measures of performance for the sieve.

12. Migration of Accidents

A lot of research has been devoted to *reducing* the number of accidents. Part of this research is focussed on identification of blackspots, their remedial treatment and the evaluation of the success of the treatment. The correction for 'regression to the mean' has become standard use.

Boyle & Wright (1984) suggest a *migration* of accidents in Greater London (UK). They studied accident data before and after remedial treatment of certain sites and investigated bordering sites as well. They concluded that, in the after period, the numbers of accidents at bordering sites increased with ca. 10% (after correction for trend). They explained this 'accident migration' into neighbouring areas by a reduced awareness of drivers as a result of the improvements: the *risk compensation hypothesis* (also see Wright & Boyle, 1987). Probably, high-risk locations tend to cluster in busy areas. Stein (1984) remarked that regression to the mean applies to locations which are treated (with an expected decrease in accident frequencies: the 'negative regression effect') as well as to locations which are not treated (with an expected *increase* in accidents, the 'positive regression effect'). McGuigan (1985a), in an analysis of personal-injury road accidents (10 011 accidents in 1980 - 1982, 70% of the accident total) in the Lothian Region (UK), shows that, from the principle of regression to the mean, an increase of 10% in accident frequencies can be expected for bordering sites, and a decrease of about 25% for treated locations. A regression effect of -31% brings the treated high-risk locations at the level of the average of high-risk locations. In the same way, the regression-to-mean effect of +15% brings the bordering untreated high-risk locations at the level of the high-risk locations in the sample (the 'inverse regression-effect', see Maher, 1987). The major part of the migration effect must be ascribed to regression to the mean.

Persaud (1987) shows that migration of risk can arise by certain conversions, e.g., by changing from two-way to multi-way stop control which is not applied to all intersections. This may lead to incorrect anticipation (drivers incorrectly expect the other stream to stop). Also, drivers may choose other routes so that traffic is redistributed, possibly because they do not want to stop or want to keep their risk-level constant. Most of the accidents saved at the converted intersections had apparently migrated to the unconverted intersections.

Maher (1987) hypothesises that the expected numbers of accidents are spatially correlated: there is a positive correlation between the expected numbers of accidents at neighboring sites, since the most important explanatory variable is exposure, or traffic flow (see e.g., Maycock & Hall, 1984). Besides this, there are other explanatory variables such as type of location and type of road, which are often comparable at neighbouring sites. It is quite natural that accident frequencies are similar at neighbouring sites. Maher presents simulation results of numbers of accidents at spatially correlated sites, computed in two different ways: (1) a spatially moving average, and (2) a clustering procedure. The contribution to the expected accident intensity (λ_j) at some 'centre' is a decreasing function of the distance between the adjacent location and the centre. Using this, Maher presents an estimator of the migration effect:

$$P_{mig} = (A_N - B_N) / B_N$$

where N means 'Neighbour' and B_N resp. A_N ('Before' resp. 'After') are the sum of observed (y_{ij}) and expected (λ_{ij}) numbers of accidents at neighbouring sites that showed more accidents than some criterion value, k . Comparable entities were determined for all treated and untreated locations (independently of neighbourhood). P_{reg} is the regression-to-the-mean effect (there was no remedial treatment in the simulations).

At a treated location, λ is large, and to be selected for improvement, y has to be large too ($y > k$, otherwise no treatment). At a *neighbouring* site, y is not large (site is *not* selected for improvement) and λ will, in general, be large, because the location borders a treated site. Since the true accident rate in the previous period was λ , the sign of $(\lambda - y) / y$, the apparent proportional increase in the accident frequencies, probably will be positive. This was simulated by Maher and they conclude that the model gives an explanation of the migration hypothesis. P_{mig} remains larger than 10%, as in the original data from London.

13. Overview of Types of Distributions and Statistics

- $f(x | \lambda)$ = $Pr(X = x | \Lambda = \lambda) = e^{-\lambda} \lambda^x / x!$; Poisson kernel
 $E(X | \lambda) = \lambda; E(X^2 | \lambda) = \lambda^2 + \lambda$
- $f_G(x)$ = $\int_0^\infty f(x | \lambda) dG(\lambda)$, marginal probability density function for x
- X_G marginal random variable whose c.d.f. is $F_G(x)$
- $F_G(x)$ marginal cumulative distribution function corresponding to $f_G(x)$
- $F_{n-1}(x)$ empirical c.d.f. of x , estimate of $F_G(x)$
such that $F_{n-1}(x) \rightarrow F_G(x)$ ($n \rightarrow \infty$) (Maritz et al., 1989, p. 13)
- $\phi_G(x)$ = $\frac{\int_0^\infty \lambda^{x+1} e^{-x} dG(\lambda)}{\int_0^\infty \lambda^x e^{-x} dG(\lambda)} = (x+1) \frac{f_G(x+1)}{f_G(x)}$; posterior mean
= regression of Λ on X_G
- $f_G(x)$ = $\int_0^\infty \frac{\lambda^x e^{-x}}{x!} dG(\lambda) = \int_0^\infty f(x | \lambda) dG(\lambda)$ mixed Poisson distribution
- $E(\lambda | X=x)$ = $\int_0^\infty \lambda f(x | \lambda) dG(\lambda) / f(x) = (x+1) \frac{f(x+1)}{f(x)}$ posterior mean for λ
= $\phi_G(x)$
- $E(\lambda f(x | \lambda))$ = $\int_0^\infty \lambda f(x | \lambda) dG(\lambda) = (x+1) f(x+1)$
- $E(\lambda^2 f(x | \lambda))$ = $\int_0^\infty \lambda^2 f(x | \lambda) dG(\lambda) = (x+1)(x+2) f(x+2)$

14. References

- Abbess, C., Jarrett, D.F., & Wright, C.C. (1981). *Accidents at blackspots: estimating the effectiveness of remedial treatment with special reference to the 'regression-to-mean effect'*. *Traff. Engng Control* 22(10), 535-542.
- Adelstein, A.M. (1951). Unpublished doctoral dissertation on the accident rates of railway operatives. Univ. of the Witwatersrand S.A.
- Aitkin, M., Anderson, D., Francis, B., & Hinde, J. (1989). *Statistical modelling in GLIM*. Oxford: Clarendon Press.
- Andreassen, D.C. & M.M. Hoque (1986). *Intersection accident frequencies*. *Traff. Engng Control*, 27(10), 514-517.
- Arbous, A.G. & Kerrich, J.E. (1951). *Accident statistics and the concept of accident proneness*. *Biometrics*, 7, 340-432.
- Asmussen, E. & Kranenburg, A. *Dynamische Systeembenadering van de Verkeersonveiligheid; Het Fasemodel van het Vervoer- en Verkeers (Onveiligheids) Proces*. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Report R-85-57.
- Bates, G.E. & Neyman, J. (1952). *Contributions to the theory of accident proneness: I. An optimistic model of the correlation between light and severe accidents*. *University of California Publications in Statistics*, 1(9), 215-254.
- Boyle, A.J. & Wright, C.C. (1984). *Accident 'migration' after remedial treatment at accident blackspots*. *Traff. Engng. Control*, 25(5), 260-267.
- Boyle, A.J. & Wright, C.C. (1985). *Accident 'migration' - or a flight of fancy?* (Letter to the Editor). *Traff. Engng. Control*, 26(7/8), 389.
- Brüde, U. & Larsson, J. (1988). *The use of prediction models for eliminating effects due to regression-to-the-mean in road accident data*. *Acc. Anal. Prev.*, 20(4), 299-310.
- Coleman, J. & Stewart, G.R. (1976). *Investigation of accident data for railroad-highway grade crossings*. *Transp. Res. Rec.* 611, 60-67.
- Efron, B. & Morris, C.N. (1977). *Stein's paradox in statistics*. *Scientific American*, 236(5), 119-128.
- Elvik, R. (1988). *Some difficulties in defining populations of 'entities' for estimating the expected number of accidents*. *Acc. Anal. Prev.*, 20(4), 261-275.
- Elvik, R. (1996). *A meta-analysis of studies concerning the safety effects of daytime running lights on cars*. *Acc. Anal. Prev.*, 28(6), 685-694.
- Farmer E. and Chambers, E.G. (1939). *A study of accident proneness among motor drivers*. I.H.R.B. Report No. 84.
- Galton, Sir F. (1877). *Typical laws of heredity*. *Proc. Roy. Inst.* 8, 282-301.
- Gardner, W., Mulvey, E.P., & Shaw, E.C. (1995). *Regression analysis of counts and rates: Poisson, Overdispersed Poisson, and Negative binomial models*. *Psych. Bull.*, 118, 392-404.
- Gipps, P.G. (1980). *Examining the safety contributions of traffic control devices*. World Conference on Urban Transport Research, London, paper E26.

Gipps, P.G. (1984). *The statistical assessment of transport data composed of counts*. Aust Rd Res., 14(2), 71-77.

Greenwood, M. & Woods, H.M. (1919). *The incidence of industrial accidents upon individuals with special reference to multiple accidents*. I.H.R.B. Report No. 4.

Hauer, E. (1980a). *Bias-by-selection: overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment*. Acc. Anal. Prev., 12, 113-117.

Hauer, E. (1980b). *Selection for treatment as a source of bias in before-and-after studies*. Traff. Engng. Control, 21(8/9), 419-421.

Hauer, E. (1981). *Evaluative Research in transport safety: The Bayesian alternative*. Publication 81-05, ISBN 0-7727-7022-0. Department of Civil Engineering, University of Toronto.

Hauer, E. (1982). *A learning disability and its cure*. Seminar on Short-term and Area-wide Evaluation of Safety Measures, Institute for Road Safety Research (SWOV), Amsterdam, April, 1982.

Hauer, E. (1986). *On the estimation of the expected number of accidents*. Acc. Anal. Prev., 18, 1-12.

Hauer, E. (1992). *Empirical Bayes approach to the estimation of 'unsafety': the multivariate regression method*. Acc. Anal. Prev., 24(5), 457-477.

Hauer, E. (1995). *On exposure and accident rate*. Traff. Engng. Control, 36(3), 134-138.

Hauer, E., Byer, P., & Joksich, H.C. (1983). *Bias-by-selection: The accuracy of an unbiased estimator*. Acc. Anal. Prev., 15, 323-328.

Hauer, E. & Lovell, J. (1986). *New directions for learning about the safety effect of measures*. Transp. Res. Rec., 1068, 96-102.

Hauer, E., Lovell, J. & Persaud, B.N. (1986). *New directions for learning about safety effectiveness*. Rep. No. FHWA/RD-86/015. Transport Safety Studies Group, Department of Civil Engineering, University of Toronto, Canada.

Hauer, E., Ng, J.C.N., & Lovell, J. (1989). *Estimation of Safety at signalized intersections*. Transp. Res. Rec., 1185, 48-61.

Hauer, E. & Persaud, B.N. (1983). *A common bias in before-and-after accident comparisons and its elimination*. Transp. Res. Rec., 905, 164-174.

Hauer, E. & Persaud, B. (1984). *The problem of identifying hazardous locations using accident data*. Transp. Res. Rec., 975, 36-42.

Hauer, E. & Persaud, B.N. (1987). *How to estimate the safety of rail-highway grade crossings and the safety effect of warning devices*. Transp. Res. Rec., 1114, 131-140.

Hauer, E., Persaud, B.N., Smiley, A., & Duncan, D. (1991). *Estimating the accident potential of an Ontario driver*. Acc. Anal. Prev., 23, 133-152.

Hauer, E., Terry, D. & Griffith, M.S. (1994). *Effect of resurfacing on safety of two-lane rural roads in New York State*. Transp. Res. Rec., 1467, 30-37.

Hutchinson, T.P. & A.J. Mayne (1977). *The year to year variability in the number of accidents*. Traff. Engng. Control, 18, 432-433.

- James, W. & Stein, C. (1961). *Estimation with quadratic loss*. Proc. Fourth Berkeley Symp. Math. Stat. Prob., 1, 361-369.
- Jarrett, D.F. (1991). *Simplified methods for assessing accident changes*. County Surveyors' Society/ Universities Transport Studies Group: Joint Seminar on the Evaluation of Road Safety Schemes.
- Jarrett, D.F. (1994). *Statistical analysis of road accident frequencies*. In: Proc. Symposium on Safety in the Environment, LNEC, Lisbon.
- Jarrett, D.F., Abbess, C.R., & Wright, C.C. (1982). *Bayesian methods applied to recent road accident blackspot studies: some recent progress*. Seminar on Short-term and Area-wide Evaluation of Safety Measures. Institute for Road Safety Research (SWOV), Amsterdam, April, 1982.
- Jarrett, D.F., Abbess, C.R., & Wright, C.C. (1988). *Empirical methods of estimating the regression-to-mean effect associated with road accident remedial treatment*. Proceedings of the International Symposium on Traffic Safety Theory and Research Methods, Institute for Road Safety Research (SWOV), Amsterdam.
- Johnson, N. & Kotz, S. (1969). *Discrete distributions*. Boston: Houghton Mifflin & Co.
- Koomstra, M.J. (1978). *Pechvogels en Brokenmakers*. In: W.A. Wagenaar, P.A. Vroon & W.H. Janssen. *Proeven op de Som: Psychonomie in het Dagelijks Leven*, 148-162. Deventer: Van Loghum Slaterus.
- Kulmala, R. (1994). *Measuring the safety effect of road measures at junctions*. *Acc. Anal. Prev.*, 26(6), 781-794.
- Lovell, J. and Hauer, E. (1986). *The safety effect of conversion to all-way stop control*. *Transp. Res. Rec.*, 1068, 103- 107.
- Maher, M.J. (1987a). *Fitting probability distributions to accident frequency data*. *Traff. Engng. Control*, 28(6), 356-357.
- Maher, M.J. (1987b). *Accident migration - a statistical explanation?* *Traff. Engng. Control*. 28(9), 480-483.
- Maher, M.J. (1990). *A bivariate negative binomial model to explain traffic accident migration*. *Acc. Anal. Prev.*, 22(5), 487-498.
- Maher, M.J. (1991). *A new bivariate negative binomial model for accident frequencies*. *Traff. Engng. Control*, 32, 422-423.
- Maher, M.J. & Mountain, L.J. (1988). *The identification of accident black-spots: A comparison of current methods*. *Acc. Anal. Prev.*, 20(2), 143-151.
- Maritz, J.S. & Lwin, T. (1989, 2nd ed.). *Empirical Bayes Methods*. London: Chapman & Hall.
- Maycock, G. & R.D. Hall. (1984). *Accidents at four-arm roundabouts*. TRRL Report LR1120, Transport and Road Research Laboratory, Crowthorne.
- McCullagh, P. & Nelder, J.A. (1989, 2nd ed.). *Generalised Linear Models*. London: Chapman & Hall.
- McGuigan, D.R.D. (1985a). *Accident migration - or a flight of fancy?* *Traff. Engng. Control*, 26(4), 229-233.
- McGuigan, D.R.D (1985b). *Reply to 'Accident migration - or a flight of fancy?'* (Letter to the Editor). *Traff. Engng. Control*, 26(7'8), 390-391.
- Miaou, S-P. (1993). *Modeling vehicle accidents and highway geometric design relationships*. *Accid. Anal. and Prev.*, 26(6), 689-709.

- Miaou, S-P. (1993). *The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions* (preprint). Transportation Research Board, 73rd Annual Meeting, January 9-13, 1994, Washington, D.C., Paper No. 940094.
- Mountain, L. Fawaz, B. & Jarrett, D. (1996). *Accident prediction models for roads with minor junctions*. *Acc. Anal. Prev.*, 28(6), 695-707.
- Nicholson, A. (1995). *Indices of accident clustering: a re-evaluation*. *Traff. Engng. Control*, 36, 291-295.
- Persaud, B. (1986). *Relating the effect of safety measures to expected number of accidents*. *Acc. Anal. Prev.*, 18(1), 63-70.
- Persaud, B.N. (1987). *'Migration' of accident risk after remedial blackspot treatment*. *Traff. Engng Control*, 28(1), 23-26 (For discussion, see *Traff. Engng. Control*, 28(1)).
- Persaud, B.N. & Hauer, E. (1984). *Comparison of two methods for debiasing before-and-after accident studies*. *Transp. Res. Rec.*, 975, p. 43-49.
- Preston, B. (1991). *Variations in the number of road casualties in different counties in England*. *Traff. Engng. Control*, 32, 302-307.
- Robbins, H. (1951). *Asymptotically subminimax solutions of compound statistical decision problems*. *Proc. Second Berkeley Symp. Math. Stat. Prob.* 131-148.
- Robbins, H. (1956). *An empirical Bayes approach to statistics*. *Proc. Third Berkeley Symp. Math. Stat. Prob.* 1, 157-163, Univ. California Press.
- Robbins, H. (1964). *The empirical Bayes approach to statistical decision problems*. (Presented as the Rietz Lecture at the 26th Annual Meeting of the Institute of Mathematical Statistics, August 29, 1963, at Ottawa, Canada. *Annual Mathematical Statistics*, 35, 1-19.
- Robbins, H. (1977). *Prediction and estimation for the compound Poisson distribution (empirical Bayes / accident proneness)*. *Proc. Natl. Acad. Sci. USA* 74(7), 2670-2671.
- Robbins, H. (1980a). *Some estimation problems for the compound Poisson distribution*. In: *Asymptotic Theory of Statistical Tests and Estimation*. New York: Academic Press, 251-257.
- Robbins, H. (1980b). *Estimation and prediction for mixtures of the exponential distribution (empirical Bayes)*. *Proc. Natl. Acad. Sci. USA* 77(5), 2382-2383.
- Robbins, H. (1980c). *An empirical Bayes estimation problem (compound Poisson distribution)*. *Proc. Natl. Acad. Sci. USA* 77(12), 6988-6989.
- Roethlisberger, F.J. and Dickson, W.J. (1949). *Management and the worker*. Harvard Univ. Press, Cambridge Univ.
- Senn, S.J. & Collie, G.S. (1988). *Accident blackspots and the bivariate negative binomial*. *Traff. Engng. Control*, 29(3) 168-169.
- Silcock, D.T., Barrell, J., & Ghee, C. (1991). *The measurement of changes in road safety*. *Traff. Engng. Control*, 32(2) 120-129.
- Stein, C. (1955). *A necessary and sufficient condition for admissibility*. *Ann. Math. Statist.*, 26, pp. 518-522.

Stein, H.S. (1984). *Accident 'migration' after remedial treatment of accident blackspots*. Letter to the Editor. *Traff. Engng. Control*, 25(12), 618.

Vernon, H.M. (1936). *Accidents and their prevention*. Cambridge Univ. Press.

Williamson, E. & Bretherton, M.H. (1963). *Tables of the negative binomial Distribution*. Wiley, 1963.

Wright, C.C. & Boyle, A.J. (1987). *Road accident causation and engineering treatment: a review of some current issues*. *Traff. Engng. Control*, 28, 475-479.

Wright, C.C., Abbess, C.R., & Jarrett, D.F. (1988). *Estimating the regression-to-mean effect associated with road accident black spot treatment: towards a more realistic approach*. *Acc. Anal. Prev.*, 20(3), 199-214.

Wyshak, G. (1974). *A program for estimating the parameters of the truncated negative binomial distribution: Algorithm A568*. *Appl. Stats*, 23(1), 87-91.