BIOMETRICS 39, 887–895
December 1983

The Univariate Generalized Waring Distribution in Relation to Accident Theory: Proneness, Spells or Contagion?

Evdokia Xekalaki[1]

Department of Statistics, University of Missouri, Columbia, Missouri 65211, U.S.A.

SUMMARY

The univariate generalized Waring distribution (UGWD) was derived by Irwin [1968, 1975, *Journal of the Royal Statistical Society, Series A* 131, 205–225 and 138, 18–31 (Part I), 204–227 (Part II), 374–384 (Part III)] as the distribution of accidents of an 'accident prone' population exposed to variable risk. This paper considers two further derivations of the UGWD in the context of accidents; these are based on a 'contagion' hypothesis and a 'spells' hypothesis, respectively. Both models assume that individuals are exposed to varying environmental risk. The problem of distinguishing between the three models is considered and some examples are given.

1. Introduction

The field of accident studies has received much attention, and various theories have been developed concerning the interpretation of the underlying factors. In 1919, Greenwood and Woods put forward three hypotheses which have formed the basis of subsequent investigations into the occurrence of accidents:

(i) *Pure chance*, which gives rise to the Poisson distribution.

(ii) *True contagion*, i.e. the hypothesis that initially all individuals have the same probability of incurring an accident but that this probability is modified by each accident sustained. This leads to what Greenwood and Woods called the 'biased distribution'.

(iii) *Apparent contagion*, i.e. the hypothesis that individuals have constant but unequal probabilities of having an accident—the resultant distribution being a compound Poisson distribution. This model is known in the literature as the 'accident proneness' model.

Under the third hypothesis, and assuming that the varying probabilities have a gamma distribution, Greenwood and Woods obtained the negative binomial distribution as the distribution of accidents. A good fit of the negative binomial distribution was then regarded as an indication of heterogeneity in the accident proneness of a whole group, until Irwin (1941) showed that this was not necessarily the case. Using a result by McKendrick (1926), he derived the negative binomial distribution for a contagion model based on the assumption that the probability of a person having an accident increases with the number of previously sustained accidents.

A fourth model that describes the occurrence of accidents rejects both the concept of accident proneness and the concept of contagion. It was formulated by Cresswell and Froggatt (1963), and is based on the assumption that each person is liable to spells, i.e. to periods of time during which the person's performance is weak. All of the person's accidents occur within those spells. The numbers of accidents within different spells are independent and are

[1] Present address: Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242, U.S.A.

Key words: Univariate generalized Waring distribution; Accident theory; Accident proneness; Accident-risk exposure; Contagion; Spells; Discrimination.

independent of the number of spells. Obviously, the negative binomial distribution can be given a 'spells' interpretation in the context of accident theory in terms of a Poisson distribution generalized by a logarithmic distribution (Kemp, 1967). Therefore, a good fit of the negative binomial is no help at all in distinguishing among the 'proneness', 'contagion' and 'spells' hypotheses. This is known as the discrimination problem between the compounded, contagion and generalized models for the negative binomial distribution and has been discussed by Arbous and Kerrich (1951), Bates and Neyman (1952), Gurland (1959) and Cane (1974, 1977). For an extensive bibliography on the accident hypotheses mentioned, see Kemp (1970).

As is well-known, in all three of these models the individuals under observation are assumed to be under equal environmental risk, a fact criticized by Irwin (1968). He suggested a three-parameter distribution which he called the 'generalized Waring distribution' (UGWD); he derived this on the basis of a hypothesis that allows separately for random factors, differences in exposure to external risk of accident, as well as for differences in proneness.

This paper demonstrates that, while the UGWD is a plausible model if accident proneness is accepted as an established fact, a satisfactory fit of this model is not to be regarded as evidence for the validity of the proneness hypothesis. Section 2 gives a brief description of Irwin's proneness model. Sections 3 and 4 retain Irwin's assumption on unequal accident-risk exposure and provide two different derivations of the UGWD in the context of the 'contagion' and 'spells' theories. Finally, in §5 the possibility of distinguishing among the three models is explored and some illustrative examples are provided.

2. Irwin's 'Proneness' Model

As mentioned above, this model assumes that a population is not homogeneous with respect to personal and environmental attributes which affect the occurrence of accidents.

Let the distribution of the number, X, of accidents for individuals of equal proneness (say ν), and of equal exposure to external risk of accident (say $\lambda \mid \nu$, i.e. λ for given ν), have probability generating function (pgf)

$$G_{X\mid\lambda,\nu}(s) = \exp\{(\lambda \mid \nu)(s-1)\}$$

in a unit time interval (0, 1). If the distributions of $\lambda \mid \nu$ and ν in the population at risk can be described by the probability density functions (pdf)

$$\{\nu^{-k}\exp(-\lambda/\nu)\lambda^{k-1}\}/\Gamma(k), \quad \nu, k > 0, \tag{1}$$

and

$$\{\Gamma(a+\rho)\nu^{a-1}(1+\nu)^{-(a+\rho)}\}/\{\Gamma(\rho)\Gamma(a)\}, \quad a, \rho > 0, \tag{2}$$

respectively, the pgf of the resulting distribution of accidents will be

$$\{\rho_{(k)}\,{}_2F_1(a, k; a+k+\rho; s)\}/(a+\rho)_{(k)},$$

i.e. the univariate generalized Waring distribution with parameters a, k and ρ which will be denoted by UGWD $(a, k; \rho)$. Here ${}_2F_1(a, b; c; z)$ denotes the Gauss hypergeometric function $\sum_{r=0}^{\infty} \{a_{(r)}b_{(r)}z^r\}/\{c_{(r)}r!\}$, where $h_{(l)} = \Gamma(h+l)/\Gamma(h)$, $h > 0$, $l \in R$. For more information about the UGWD the reader is referred to the work of Irwin (1963, 1968, 1975), Xekalaki (1981) and the references therein, and Xekalaki (1983).

3. The 'Contagion' Model

In this section the assumptions of the classical contagion model developed by Greenwood and Yule (1920) are extended by considering a population of individuals exposed to varying accident risk.

Assume that at time $t = 0$ none of the individuals has had an accident. This will be true if we are concerned, for example, with a population who are just beginning a new type of work. Suppose that during the time period from t to $t + dt$ a person who has had x accidents by Time t can have another accident with a probability of $\{(k + x)/(1 + \lambda t)\}\lambda dt$ (independent of the times of the previous accidents), where k is a positive constant and λ refers to the individual's risk exposure. At $t = 0$, since $x = 0$ the probability of an accident is $k\lambda dt$. Hence what the model basically assumes is that, initially, the probability of having an accident is not the same for each individual but depends on the external conditions; later, the probability is also affected by the number of preceding accidents. Under these assumptions and if differences in the exposure to accident risk can be thought of as governed by a distribution with pdf given by (2), the final distribution of accidents over a unit period of time turns out to be UGWD(a, k; ρ). So two different hypotheses lead to the same distribution.

4. The 'Spells' Model

Let us now consider a variant of the 'spells' model due to Cresswell and Froggatt (1963), that rejects the presence of proneness and contagion.

Assume that every individual is liable to spells and that the number of spells in a given time period $(0, t)$ is a Poisson variable with parameter θt, $\theta > 0$. Suppose that no accidents can occur outside spells and that the probability of an accident occurring within a spell is dependent on the risk exposure of the particular individual. In particular, suppose that within a spell a person can have

$$\left.\begin{array}{l} 0 \text{ accidents with probability } 1 - m\log(1+\lambda) \\ \text{or} \\ n \text{ accidents } (n \geqslant 1) \text{ with probability } m\{\lambda/(1+\lambda)\}^n/n, \quad m > 0,\ \lambda > 0, \end{array}\right\} \quad (3)$$

where λ is the external risk parameter for the given individual. Assume further that the numbers of accidents arising out of different spells are independent and are also independent of the number of spells. Then, if differences in the risk exposure can be described by a distribution with pdf given by (2), the resulting accident distribution will have pgf

$$\{\rho_{(a)}{}_2F_1(a, \theta mt; a + \theta mt + \rho; s)\}/(\rho + \theta mt)_{(a)}.$$

Hence, in a unit time period the distribution of accidents is UGWD(a, θm; ρ). So, as seen from §§2, 3 and 4, three completely different sets of hypotheses give rise to exactly the same form of distribution.

It is worth noticing that the form of the distribution of λ in the models of §§3 and 4 is more general than that considered in §2. It is, however, a reasonable choice as it implies a beta distribution of the first kind (Pearson Type I) for the parameter $q = \lambda/(1 + \lambda)$ of the negative binomial distribution of $X|\lambda$.

5. Deciding about the Underlying Model

Statisticians have always been tempted to look for ways of discriminating among different models that give rise to the same distribution. With regard to the negative binomial distribution as an accident distribution, most attempts seem to have been concentrated on distinguishing between the proneness and contagion models. The papers by Bates and Neyman (1952) and Bates (1955) cover part of the work that has been done on the subject, although they particularly focus on distinguishing between different forms of contagion. Shaw and Sichel (1971) tried to prove or disprove proneness by ranking individuals on an accident performance scale based on their average interval between successive accidents.

Cane's (1974) paper seems to be the first instance in the literature of a systematic attempt to discriminate between the proneness and contagion models of the negative binomial distribution. The basic idea presented by Cane was that one might be able to decide between the two possibilities if complete information about the accidents were available. What is meant by complete information is best described in the author's words: 'by this I mean that the time of each accident for each person in the sample is known'. She showed, however, that one cannot distinguish between the two models, even with complete information. In particular, she showed that, for both models, the conditional distribution of the times, t_i, $i = 1, 2, \ldots, n$, at which accidents have occurred in a time period from 0 to T is the same, namely that of an ordered sample from a uniform distribution on $(0, T)$ with pdf $n!\, T^{-n}$. In fact, this is the case for any compound Poisson accident distribution whose compounding distribution has finite moments (Cane, 1977); a corresponding process, indistinguishable from it, can always be defined to describe the accident experience of the population.

Suppose now that we have the detailed records of the accidents suffered by individuals subject to varying risks and that the observed distribution of the total number of accidents is UGWD$(a, k; \rho)$. Then, following Cane (1974), we have that, for individuals exposed to different accident risks within the time interval (0, 1), the probability of n accidents at times t_i, $i = 1, 2, \ldots, n$, $t_0 \equiv 0 < t_1 < \cdots < t_n < 1$, is

$$\int_0^\infty \exp\{-(\lambda \mid \nu)(1 - t_n)\} \prod_{i=1}^{n} \exp\{-(\lambda \mid \nu)(t_i - t_{i-1})\}(\lambda \mid \nu)\, dt_i dF(\lambda \mid \nu)$$

in the case of the proneness model, and is

$$\int_0^\infty \left(\frac{1 + \lambda t_n}{1 + \lambda}\right)^{k+n} \prod_{i=1}^{n} \left(\frac{1 + \lambda t_{i-1}}{1 + \lambda t_i}\right)^{k+i-1} \left(\frac{k + i - 1}{1 + \lambda t_i}\right)\lambda\, dt_i\, dH(\lambda)$$

in the case of the contagion model. Here, $F(\lambda \mid \nu)$ and $H(\lambda)$ represent the distribution functions of the distributions defined by the densities (1) and (2), respectively. On integration the above expressions reduce to

$$(n!\, dt_1 \ldots dt_n)\left(\frac{k_{(n)}}{n!}\right)\left(\frac{\nu}{1 + \nu}\right)^n\left(\frac{1}{1 + \nu}\right)^k$$

and

$$(n!\, dt_1 \ldots dt_n)\left\{\frac{\rho_{(k)}}{(a + \rho)_{(k)}}\right\}\left\{\frac{a_{(n)}k_{(n)}}{(a + k + \rho)_{(n)}}\right\}\left(\frac{1}{n!}\right),$$

respectively. Hence, in both cases the t_i, $i = 1, 2, \ldots, n$, conditional on n accidents in the time interval (0, 1), have a joint pdf given by

$$f(t_1, \ldots, t_n) = n!, \tag{4}$$

i.e. they form an ordered sample of size n from a uniform distribution on (0, 1). This implies that the availability of information on the times of the occurrence of accidents is not sufficient to guide one's choice between the proneness and contagion models.

Let us now consider the problem of finding the joint distribution of t_i, $i = 1, 2, \ldots, n$, for individuals who have n accidents in a unit period of time under the spells model. For fixed λ, accidents occur as events in a generalized Poisson process:

$$X(t) = \sum_{i=1}^{N(t)} Y_i, \quad N(t) \sim \text{Poisson}(\theta t),$$

where $\theta > 0$, $t \geq 0$ and Y_i i.i.d. with probability function given by (3). Consequently, the required probability is

$$\int_0^{\infty} (1+\lambda)^{-\theta m(1-t_n)} \left[\prod_{i=1}^{n} \{\lambda\theta m(1+\lambda)^{-\theta m(t_i - t_{i-1}) - 1}\, dt_i\} \right] dH(\lambda),$$

where $H(\lambda)$ is defined as before, i.e. the probability is

$$\left\{ \frac{(\theta m)^n \rho_{(a)} a_{(n)}}{(\theta m + \rho)_{(a+n)}} \right\} dt_1 \ldots dt_n.$$

Hence, conditional on n accidents during a time period from 0 to 1, the joint pdf of t_i, $i = 1, 2, \ldots, n$, is

$$n!(\theta m)^n / (\theta m)_{(n)}. \tag{5}$$

The form of this differs from that arising under the proneness and contagion models. This fact in itself is very interesting as far as establishing the presence of spells is concerned, as it implies the following: if an observed accident distribution of the UGWD type has arisen from the spells model, the time intervals $(0, t_i)$, $i = 1, 2, \ldots, n$, given a total of n accidents, will be jointly distributed with pdf given by (5). Any departure from this distribution is, then, evidence against the spells model.

Of course, if on the available evidence one has to reject (5) in favor of (4), then one is faced again with the question: 'proneness or contagion?' This cannot be answered by studying the distribution of t_i. In such cases, a promising way to tackle the problem would appear to be to study the distribution of ν among persons who have had x accidents in the period 0 to 1. This may enable one to 'estimate' a person's proneness on the basis of the incurred number of accidents.

It can be shown that the pdf of $\nu^* \equiv \nu \mid (X = x)$ is given by

$$\left\{ \frac{\Gamma(a+k+\rho+x)}{\Gamma(\rho+k)\Gamma(a+x)} \right\} \nu^{*x+a-1} (1+\nu^*)^{-(a+k+\rho+x)}.$$

This implies that $\{(k+\rho)/(a+x)\}\nu^*$ has the F distribution with $2(a+x)$ and $2(k+\rho)$ degrees of freedom and hence, if $f_{1-\frac{1}{2}\alpha}$ denotes the $100(1-\frac{1}{2}\alpha)$ percentile of this distribution, $[\{(a+x)/(k+\rho)\}f_{\frac{1}{2}\alpha}, \{(a+x)/(k+\rho)\}f_{1-\frac{1}{2}\alpha}]$ is a $100(1-\alpha)\%$ confidence interval for $\nu \mid (X = x)$, $x = 0, 1, 2, \ldots$. This approach is illustrated in what follows.

Table 1 presents the distribution of accidents that occurred to 414 machinists in a

Table 1
Distribution of accidents to 414 machinists over a period of three months (Greenwood and Yule, 1920)

Number of accidents	Observed frequency		Expected frequency	
0	296		292.951	
1	74		77.701	
2	26		25.461	
3	8		9.683	
4	4	10	4.119	8.204
5	4		1.912	
6	1		0.953	
7	0		0.503	
8	1		0.717	
Total	414		414.000	

$\hat{a} = 2.86 \pm 0.93$, $k = 1$, $\hat{\rho} = 6.91 \pm 1.95$, $\chi^2 = 0.9054$, df $= 2$, $\text{pr}(\chi_2^2 \geq 0.9054) = .65$

Table 2
90% confidence limits for $\nu \mid (X = x)$, *corresponding to* $\hat{a} = 2.86$, $k = 1$, $\hat{\rho} = 6.91$ *for accidents to machinists*

x	Lower limit	Upper limit
0	0.092	0.990
1	0.152	1.263
2	0.217	1.530
3	0.285	1.795
4	0.355	2.057
5	0.426	2.317
6	0.498	2.576
7	0.570	2.835
8	0.644	3.095

period of three months (Greenwood and Yule, 1920). To this observed distribution, UGWD(a, 1; ρ) was fitted by equating the first two factorial moments to their observed values. The UGWD fits the data satisfactorily ($P = .65$). From Table 2 one cannot really establish that proneness is the underlying factor. There is a great deal of overlap between the confidence intervals for different values of x so one, for example, cannot assert that persons with eight accidents are more prone than persons with three accidents. One would, therefore, be tempted to conclude that proneness does not seem to be present in this particular accident situation. However, as demonstrated below, this can be a dangerous conclusion as there are cases where, even when the presence of proneness has been established from prior information on the group of people under investigation, the confidence intervals for $\nu \mid (X = x)$ may indicate otherwise.

This seems to be the case with the distribution of accidents in which workers in a soap factory were involved over a five-month period; see Table 3 (Newbold, 1927). Irwin (1975)

Table 3
Distribution of accidents to 447 men in a soap factory over a period of five months (*Irwin*, 1975)

Number of accidents	Observed frequency		Expected frequency
0	239		237
1	98		108
2	57		50
3	33		24
4	9		12
5	2		7
6	2		4
7	1	7	5
8	0		
9	4		
10	1		
11	0		
12	0		
13	1		
Total	447		447

$\hat{a} = 4.31 \pm 1.61$, $\hat{k} = 1.33 \pm 0.81$, $\hat{\rho} = 6.92 \pm 2.23$, $\chi^2 = 10.6$, df $= 4$, pr$(\chi_4^2 \geq 10.6) = .032$

Table 4
90% confidence limits for $\nu|(X = x)$, corresponding to $\hat{a} = 4.31$, $\hat{k} = 1.33$, $\hat{\rho} = 6.92$ for accidents to soap factory workers

x	Lower limit	Upper limit
0	0.163	1.354
1	0.228	1.605
2	0.294	1.855
3	0.363	2.103
4	0.432	2.351
5	0.502	2.597
6	0.572	2.844
7	0.643	3.092
8	0.715	3.335
9	0.786	3.584
10	0.858	3.829
11	0.930	4.071
12	1.003	4.320
13	1.075	4.567

fitted these data with the UGWD by the method of maximum likelihood on his accident proneness hypothesis. Newbold (1927) found a correlation of 0.36 between the number of accidents, X, and the various departments in the factory; this indicated little (if any) variation in the external risk. Hence, as Irwin (1968) demonstrated, any heterogeneity in the population reflected mostly proneness. However, from Table 4 one cannot obviously assert the presence of proneness. One should probably note at this point that the UGWD is symmetrical in a and k. So in each of the examples mentioned it is possible to obtain a second set of confidence intervals for $\nu|(X = x)$ by reversing the roles of a and k. But, again any attempt to establish proneness on the basis of these confidence intervals (Tables 5 and 6) proves inconclusive. It seems, therefore, that 'estimating' proneness conditional on the number of incurred accidents is not a promising approach.

It appears from the foregoing analysis that the problem of discriminating among the three models that give rise to the UGWD has not been solved completely. Some light has been shed, however, as it seems possible to determine whether the spells model is the underlying

Table 5
90% confidence limits for $\nu|(X = x)$, corresponding to $a = 1$, $\hat{k} = 2.86$, $\hat{\rho} = 6.91$ for accidents to machinists

x	Lower limit	Upper limit
0	0.005	0.224
1	0.035	0.715
2	0.079	0.880
3	0.130	1.064
4	0.185	1.252
5	0.242	1.442
6	0.200	1.983
7	0.343	1.822
8	0.405	2.012

Table 6
90% confidence limits for $\nu \mid (X = x)$, *corresponding to* $\hat{a} = 1.33$, $\hat{k} = 4.31$, $\hat{\rho} = 6.92$ *for accidents to soap factory workers*

x	Lower limit	Upper limit
0	0.006	0.408
1	0.036	0.584
2	0.077	0.756
3	0.123	0.924
4	0.172	1.090
5	0.223	1.255
6	0.276	1.418
7	0.329	1.581
8	0.383	1.743
9	0.437	1.905
10	0.492	2.068
11	0.548	2.227
12	0.603	2.390
13	0.659	2.551

model or not, provided complete records of an accident situation are available. Otherwise, distinguishing between proneness and contagion does not seem to be feasible with the current statistical approaches.

ACKNOWLEDGEMENTS

I would like to thank Professor C. D. Kemp for many discussions on the subject of accident proneness, the referees for some helpful comments and the Alexandros Onassis Foundation for a grant in support of this study. This paper is dedicated to the memory of Dr J. O. Irwin.

RÉSUMÉ

La distribution univariée de Waring généralisée (UGWD) fut introduite par Irwin [1968, 1975, *Journal of the Royal Statistical Society, Series A* 131, 205–225 et 138, 18–31 (Partie I), 204–227 (Partie II), 374–384 (Partie III)] comme la distribution des accidents dans une population 'prédisposée' soumise à des risques variables. Cet article présente deux prolongements de la UGWD pour l'étude des accidents basés soit sur une hypothèse de contagion, soit sur une hypothèse de relais. Ces deux modèles supposent que les individus sont soumis à des risques environnementaux variables. On considère alors le problème de la distinction entre ces troix modèles et on donne quelques exemples.

REFERENCES

Arbous, A. G. and Kerrich, J. E. (1951). Accident statistics and the concept of accident proneness. *Biometrics* 7, 340–432.

Bates, G. E. (1955). Joint distributions of time intervals for the occurrence of successive accidents in a generalized Polya scheme. *Annals of Mathematical Statistics* 26, 705–720.

Bates, G. E. and Neyman, J. (1952). Contributions to the theory of accident proneness. *University of California Publications in Statistics* 1, 215–275.

Cane, V. R. (1974). The concept of accident proneness. *Bulletin of the Institute of Mathematics, Bulgarian Academy of Sciences* 15, 183–188.

Cane, V. R. (1977). A class of non-identifiable stochastic models. *Journal of Applied Probability* 14, 475–482.

Cresswell, W. L. and Froggatt, P. (1963). *The Causation of Bus Driver Accidents.* London: Oxford University Press.

Greenwood, M. and Woods, H. M. (1919). *On the Incidence of Industrial Accidents upon Individuals,* *with Special Reference to Multiple Accidents.* Report of the Industrial Fatigue Research Board, No. 4, 1–28. London: His Majesty's Stationery Office.

Greenwood, M. and Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attack of disease or repeated accidents. *Journal of the Royal Statistical Society* **83**, 255–279.

Gurland, J. (1959). Some applications of the negative binomial and other contagious distributions. *American Journal of Public Health* **49**, 1388–1399.

Irwin, J. O. (1941). Discussion on Chambers and Yule's paper. *Journal of the Royal Statistical Society, Supplement* **7**, 101–109.

Irwin, J. O. (1963). The place of mathematics in medical and biological statistics. *Journal of the Royal Statistical Society, Series A* **126**, 1–44.

Irwin, J. O. (1968). The generalized Waring distribution applied to accident theory. *Journal of the Royal Statistical Society, Series A* **131**, 205–225.

Irwin, J. O. (1975). The generalized Waring distribution. *Journal of the Royal Statistical Society, Series A* **138**, 18–31 (Part I), 204–227 (Part II), 37--384 (Part III).

Kemp, C. D. (1967). On a contagious distribution suggested for accident data. *Biometrics* **23**, 241–255.

Kemp, C. D. (1970). 'Accident proneness' and discrete distribution theory. In *Random Counts in Scientific Work, Vol.* 2, G. P. Patil (ed.), 41–65. State College: Pennsylvania State University Press.

McKendrick, A. G. (1926). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society* **44**, 98–130.

Newbold, E. M. (1927). Practical applications of the statistics of repeated events, particularly to industrial accidents. *Journal of the Royal Statistical Society* **90**, 487–547.

Shaw, L. and Sichel, H. S. (1971). *Accident Proneness.* Oxford: Pergamon Press.

Xekalaki, E. (1981). Chance mechanisms for the univariate generalized Waring distribution and related characterizations. In *Statistical Distributions in Scientific Work, Vol.* 4 (*Models, Structures and Characterizations*), C. Taillie, G. P. Patil and B. Baldessari (eds), 157–171. Dordrecht: Reidel.

Xekalaki, E. (1983). Infinite divisibility, completeness and regression properties of the univariate generalized Waring distribution. *Annals of the Institute of Statistical Mathematics, Part A* **35**, 161–171.

Received December 1980; *revised May and October* 1982