International Journal of Statistics and Applied Mathematics 2018; 3(2): 41-44

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2018; 3(2): 41-44

www.mathsjournal.com
Received: 16-01-2018
Accepted: 17-02-2018

Dr. SG Parekh
Faculty of Business Administration, Dharmsinh Desai University, Nadiad, Gujarat, India

SR Patel
Department of Statistics, Sardar Patel University, Vallabh Vidyanagar, India

Correspondence
Dr. SG Parekh
Faculty of Business Administration, Dharmsinh Desai University, Nadiad, Gujarat, India

A study of compound Poisson and compound negative binomial frailty models

SG Parekh and SR Patel

Abstract
In this paper we have introduced the Compound Poisson Frailty model and Compound Negative Binomial Frailty Model and we have tried to get maximum likelihood estimators. We developed the distribution by use of Laplace transform. The likelihood equations have intrinsic in nature.

Keywords: Compound Poisson distribution, Compound Negative Binomial distribution, Frailty model, Laplace transformation, Maximum likelihood estimators.

1. Introduction
To properly interpret result of survival analysis, we have to consider the fact that individual risks differ in, possibly, unknown ways. This heterogeneity may be difficult to assess, but it nevertheless of great importance. It may distort observed survival curves and hazard rates. This has been discussed by many authors e.g., Manton et.al, (1986) [9], Vaupel and Yashin (1985) [11] and Hougaard (1984, 1986a, 1986b) [4, 5, 6].
When intensities are estimated, for instance by incidence rates, one may be interested in how they change as a function of time. Quite often they are seen to be rising at the start, reaching a maximum and then declining. Hence the intensity has a unimodel shape with a finite mode.
Aalen (1988, 1992) [1, 2] introduced the compound Poisson distribution as a frailty distribution. This model yields a subgroup of zero frailty, which survives forever. In spite of the fact that the density of the continuous part is only given as an infinite series which has to be calculated numerically, the distribution is mathematically convenient. It may also be seen as a natural choice. Similarly Compound Negative Binomial frailty Model is also useful in demography and medicine. Here also two parts have been studied separately.
In section 2 we study the Compound Poisson frailty model with the use of Laplace transform. In section 3 estimation of Compound Poisson frailty model has been done. In section 4 Compound Negative Binomial frailty model has been introduced. Section 5 discusses estimation of Compound Negative Binomial frailty model.

2. Compound Poisson frailty model
The compound Poisson distribution can be constructed as the sum of a Poisson distributed number of independent and identical gamma distributed random variables. When each individual experiences a random number of hits, each of a random size, this distribution can be viewed as a hit model.

$$Z = \begin{cases} X_1 + X_2 + \cdots + X_N : if\ N > 0, \\ 0 : if\ N = 0 \end{cases} \qquad (2.1)$$

where N is Poisson distributed with parameter ρ, where ρ is expected value of N, while $X_1, X_2, \ldots$ are independent and gamma distributed with $X_i \sim G(k, \lambda)$. The Laplace transforms of gamma and Poisson distributions are given by

$$L_x(s) = \left(1 + \frac{s}{\lambda}\right)^{-k}$$

and $\quad L_N(s) = e^{-\rho+\rho e^{-s}}$ respectively.

Now, $\quad N \sim P(\rho)\ P(N=n) = \frac{e^{-\rho}\rho^n}{n!}$

$$L_N(s) = Ee^{-SN}$$
$$L_N(s) = \sum_{n=0}^{\infty} e^{-sn}.\frac{e^{-\rho}\rho^n}{n!}$$
$$= e^{-\rho}\sum_{n=0}^{\infty}\frac{(e^{-s}\rho)^n}{n!}$$
$$= e^{-\rho}e^{\rho e^{-s}}$$
$$L_N(s) = e^{-\rho+\rho e^{-s}}$$
$$L(s) = Ee^{-SZ}, \quad Z = X_1+X_2+\cdots+X_N$$
$$= Ee^{-S(X_1+X_2+\cdots+X_N)}$$
$$= E\left(L_X(s)\right)^N$$
$$= Ee^{NlogL_X(s)}$$
$$= L_N\left(-logL_X(s)\right) \qquad (2.2)$$

Using (2.2), we have

$$L(s) = L_N\left(-logL_X(s)\right)$$
$$L(s) = e^{-\rho+\rho e^{log\left(L_X(s)\right)^{-1}}}$$
$$= e^{-\rho+\rho\left(L_X(s)\right)^{-1}}$$
$$= e^{-\rho+\rho\left(1+\frac{s}{\lambda}\right)^{-k}}$$

Let us make parameterization of ρ, λ , k as $k = -c, \lambda = \lambda$ and $\rho = -\left(\frac{k\lambda^c}{c}\right) = \lambda^c$

$$L(s) = e^{-\lambda^c+\lambda^c\left(1+\frac{s}{\lambda}\right)^{-k}}$$
$$= e^{-\lambda^c+\lambda^c\left(\frac{\lambda+s}{\lambda}\right)^{-k}}$$
$$= e^{-\lambda^c+\lambda^c(\lambda+s)^c\lambda^{-c}}$$
$$= e^{-\lambda^c+(\lambda+s)^c}$$
$$= e^{-\frac{k}{c}[(\lambda+s)^c-\lambda^c]} \qquad (2.3)$$

Parameter c divides the class of distributions in two major parts. The distribution is a power variance function (PVF) distribution, for $c \geq 0$. Aalen (1988)[1] has suggested the extension to $c < 0$ and yields the compound Poisson distribution. The two parts are separated by the gamma distribution (c = 0). The different parameterization was used by Aalen. The compound Poisson distribution is denoted by $CP(\gamma, k, \lambda)$.

The marginal survival and hazard function in case of a compound Poisson frailty model using the Laplace transform given in (2.3) are

$$S(t) = L\left(H_0(t)\right)$$
$$\therefore\ S(t) = e^{-\frac{k}{c}\left[\left(\lambda+H_0(t)\right)^c-\lambda^c\right]}$$

and
$$h(t) = -\frac{d}{dt}\left(logS(t)\right)$$
$$= -\frac{d}{dt}\left\{-\frac{k}{c}\left[\left(\lambda+H_0(t)\right)^c-\lambda^c\right]\right\}$$
$$= \frac{k}{c}.c\left(\lambda+H_0(t)\right)^{c-1}$$
$$= kh_0(t)\left(\lambda+H_0(t)\right)^{c-1}$$

Further, Using $EZ = 1$ and $\sigma^2 = \frac{1-c}{\lambda}\ \therefore\ \lambda = \frac{1-c}{\sigma^2}$

$$S(t) = e^{-\frac{k}{c}\left[\left(\lambda+H_0(t)\right)^c-\lambda^c\right]}$$
$$= e^{-\frac{k}{c}\left[\left(\frac{1-c}{\sigma^2}+H_0(t)\right)^c-\left(\frac{1-c}{\sigma^2}\right)^c\right]}$$
$$= e^{-\frac{k}{c}\left[\left(\frac{1-c}{\sigma^2}\right)^c\left(1+\frac{\sigma^2}{1-c}H_0(t)\right)^c-\left(\frac{1-c}{\sigma^2}\right)^c\right]}$$
$$= e^{-\frac{k}{c}\left(\frac{1-c}{\sigma^2}\right)^c\left[\left(1+\frac{\sigma^2}{1-c}H_0(t)\right)^c-1\right]}$$
$$= e^{\left(\frac{1-c}{\sigma^2}\right)^c\left[\left(1+\frac{\sigma^2}{1-c}H_0(t)\right)^c-1\right]} \quad \because k = -c$$

and
$$h(t) = kh_0(t)\left(\lambda+H_0(t)\right)^{c-1}$$
$$= kh_0(t)\left(\frac{1-c}{\sigma^2}+H_0(t)\right)^{c-1}$$
$$= kh_0(t)\ \left(\frac{1-c}{\sigma^2}\right)^{c-1}\left(1+\frac{\sigma^2}{1-c}\ H_0(t)\right)^{c-1}$$
$$h(t) = \frac{kh_0(t)\left(\frac{1-c}{\sigma^2}\right)^{c-1}}{\left(1+\frac{\sigma^2}{1-c}H_0(t)\right)^{1-c}},$$

It should be noted that when $c < 0$ the integral of $h(t)$ over $[0,\infty)$ is finite. Also note that the survival function is incomplete because a fraction of individuals has zero frailty that will never experience the event under study. Aalen (1992)[2] has used model to model the incidence of marriage of women born in Denmark. Marriage is an example of an event that does not happen to everybody. A certain percentage of individuals never marry and models of marriage incidence have to account for this. The fact that the observed incidence peaks around age 23 and becomes rather low after age 30 is therefore interpreted to be a selection phenomenon due to heterogeneity, meaning that those who are most prone to marriage will marry quite early, and the reminder will have a lower tendency to marry. A second application of Aalen (1992)[2] deals with fertility data in Norwegian woman. It is known that around 5% to 10% of all couples are unable to conceive children. Naturally, P(Z = 0) (where Z denotes the frailty to conceive a child) will be the probability of infertility, and the variation of the frailty variable Z expresses the varying fecund abilities among fertile couples.

The model was applied by Hougaard et al. (1994)[4] to diabetic nephropathy onset data, a serious complication experienced by some diabetic patients. This model was also applied by Aalen and Tretli (1999)[3] to testicular cancer. Testicular cancer has two striking epidemiological features. First, its incidence has increased speedily over the past few decades and secondly, the incidence is greatest among younger men, and then declines from a certain age. The idea of the model is that a subgroup of men is particularly susceptible to testicular cancer, which results in selection over time. The model is fit to incidence data from the Norwegian Cancer Registry collected between 1953 and 1993. This work is extended by Moger et al. (2004a)[10]. Haukka et al. (2003)[8] applied the model to schizophrenia data from the Finnish population born 1950 - 1968. They concluded that only a small part of the population is susceptible to schizophrenia and found increasing individual risk with higher age among the vulnerable part of the population.

3. Estimation of Compound Poisson distribution

After some simulation the p.d.f. of Compound Poisson variate z is

$$f(z;\gamma,\nu,\rho) = \begin{cases} exp[-(\rho+\nu z)]\frac{1}{z}\sum_{n=1}^{\infty}\frac{\rho^n(\nu z)^{n\gamma}}{\Gamma(n\gamma)n!}; z > 0, \rho > 0 \\ 0 : o.w. \end{cases} \qquad (3.1)$$

Consider n observations $z_1, z_2, \ldots, z_n$ from (3.1) so that the likelihood will be

$$L = \prod_{i=1}^{n}\frac{e^{-(\rho+\nu z_i)}}{z_i}\prod_{i=1}^{n}\sum_{k=1}^{\infty}\frac{\rho^k(\nu z_i)^{k\gamma}}{\Gamma(k\gamma)k!}$$

so that

$$logL = -\sum_{i=1}^{n}(\rho + \nu z_i) - \sum_{i=1}^{n} logz_i + \sum_{i=1}^{n} log\left\{\sum_{k=1}^{\infty}\frac{\rho^k (\nu z_i)^{k\gamma}}{\Gamma(k\gamma)k!}\right\}$$

After differentiating $logL$ w.r.t. ρ, ν and γ we have

$$\frac{\partial logL}{\partial \rho} = \sum_{i=1}^{n}\frac{\sum_{k=1}^{\infty}\frac{k\rho^{k-1}(\nu z_i)^{k\gamma}}{\Gamma(k\gamma)k!}}{\sum_{k=1}^{\infty}\frac{\rho^k(\nu z_i)^{k\gamma}}{\Gamma(k\gamma)k!}} = n \qquad (3.2)$$

$$\frac{\partial logL}{\partial \nu} = \sum_{i=1}^{n}\frac{\sum_{k=1}^{\infty}\frac{k\gamma\rho^{k}(\nu z_i)^{k\gamma-1}z_i}{\Gamma(k\gamma)k!}}{\sum_{k=1}^{\infty}\frac{\rho^k(\nu z_i)^{k\gamma}}{\Gamma(k\gamma)k!}} = \sum_{i=1}^{n} z_i \qquad (3.3)$$

$$\frac{\partial logL}{\partial \gamma} = \sum_{i=1}^{n}\frac{\sum_{k=1}^{\infty}\frac{\rho^k}{k!}\left\{\frac{\Gamma(k\gamma)(\nu z_i)^{k\gamma}log(\nu z_i) - (\nu z_i)^{k\gamma}\frac{\partial\Gamma(k\gamma)}{\partial\gamma}}{(\Gamma(k\gamma))^2}\right\}}{\sum_{k=1}^{\infty}\frac{\rho^k(\nu z_i)^{k\gamma}}{\Gamma(k\gamma)k!}} = 0 \qquad (3.4)$$

The asymptotic equations with $O(k^{-2})$ after some simplifications, we shall get following equations.

$$\frac{n\Gamma(2\gamma)}{n\Gamma(2\gamma) - \Gamma(\gamma)\nu^\gamma\sum_{i=1}^{n} {z_i}^{\gamma}} = \rho \qquad (3.5)$$

$$\frac{\Gamma(2\gamma)}{\rho\Gamma(\gamma)\nu^\gamma}\left\{\frac{\nu}{\gamma}\sum_{i=1}^{n} z_i - n\right\} = \sum_{i=1}^{n} {z_i}^{\gamma} \qquad (3.6)$$

$$log\nu + log\tilde{z} - \frac{\partial log\Gamma(\gamma)}{\partial\gamma} = 0 \qquad (3.7)$$

Where $\tilde{z} = (\prod_{i=1}^{n} z_i)^{\frac{1}{n}}$

4. Compound negative binomial distribution

A lot of items has been inspected till one gets two items defective, it follows negative binomial distribution. Persons go on changing the jobs until he gets a suitable job which is an example of geometric distribution which is a part of negative binomial distribution. Thus N follows either geometric or negative binomial distribution and if X_i follows gamma distribution,

Let $Z = X_1 + X_2 + \cdots + X_N$

Then Z follows compound negative binomial distribution. N has mass function as

$$P(N = n) = \begin{cases}\binom{N+r+1}{N} p^r q^N;\ N = 0,1,2,\ldots;\ 0 < p < 1;\ q = 1-p \\ 0;\ otherwise\end{cases}$$

Then

where $r = 1,2,\ldots$ and $X_1, X_2, \ldots$, are independent gamma variates with scale parameter ν and shape parameter γ having density function.

$$f(X) = \begin{cases}\frac{\nu^\gamma}{\Gamma(\gamma)} X^{\gamma-1}e^{-\nu X};\ X > 0, \nu > 0, \gamma > 0 \\ 0;\ otherwise\end{cases}$$

Z has got distribution in two parts a discrete part with $\mathrm{P}(Z = 0) = p^r$ and the continuous part has p.d.f. as

$$f(z) = \begin{cases}p^r \frac{1}{z} e^{-\nu z}\sum_{N=1}^{\infty}\binom{N+r-1}{N} q^N \frac{(\nu z)^{N\gamma}}{\Gamma(N\gamma)};\ z > 0, \nu > 0, \gamma > 0 \\ 0;\ otherwise\end{cases} \qquad (4.1)$$

where r = 1, 2, …, $0 < p < 1;\ q = 1 - p$,

5. Estimation of Compound Negative Binomial distribution

After some simplification the p.d.f. of Compound Negative Binomial variable z is

$$f(z) = \begin{cases}p^r \frac{1}{z} e^{-\nu z}\sum_{N=1}^{\infty}\binom{N+r-1}{N} q^N \frac{(\nu z)^{N\gamma}}{\Gamma(N\gamma)};\ z > 0, \nu > 0, \gamma > 0 \\ 0;\ otherwise\end{cases} \qquad (5.1)$$

Taking n observations $z_1, z_2, \ldots, z_n$ from (5.13.1), the likelihood L will be

$$L = p^{rn}\prod_{i=1}^{n}\left(\frac{1}{z_i}\right)e^{-\nu\sum z_i}\prod_{i=1}^{n}\sum_{N=1}^{\infty}\binom{N+r-1}{N} q^N \frac{(\nu z_i)^{N\gamma}}{\Gamma(N\gamma)}$$

and

$$\log L = rnlogp - \sum_{i=1}^{n} logz_i - \nu\sum_{i=1}^{n} z_i + \sum_{i=1}^{n} log\left\{\sum_{N=1}^{\infty}\binom{N+r-1}{N} q^N \frac{(\nu z_i)^{N\gamma}}{\Gamma(N\gamma)}\right\}$$

Differentiating $log\ L$ with respect to parameters, the m.l.e. will be

$$\frac{\partial logL}{\partial p} = \sum_{i=1}^{n}\frac{\sum_{N=1}^{\infty}\binom{N+r-1}{N}\frac{(\nu z_i)^{N\gamma}}{\Gamma(N\gamma)}q^{N-1}}{\sum_{N=1}^{\infty}\binom{N+r-1}{N}q^N\frac{(\nu z_i)^{N\gamma}}{\Gamma(N\gamma)}} = \frac{rn}{p} \qquad (5.2)$$

$$\frac{\partial logL}{\partial \nu} = \sum_{i=1}^{n}\frac{\sum_{N=1}^{\infty}\binom{N+r-1}{N}q^N\frac{N\gamma(\nu z_i)^{N\gamma-1}z_i}{\Gamma(N\gamma)}}{\sum_{N=1}^{\infty}\binom{N+r-1}{N}q^N\frac{(\nu z_i)^{N\gamma}}{\Gamma(N\gamma)}} = \sum_{i=1}^{n} z_i \qquad (5.3)$$

$$\frac{\partial logL}{\partial \gamma} = \sum_{i=1}^{n}\frac{\sum_{N=1}^{\infty}\binom{N+r-1}{N}q^N\frac{\Gamma(N\gamma)(\nu z_i)^{N\gamma}log(\nu z_i).N - (\nu z_i)^{N\gamma}\frac{\partial\Gamma(N\gamma)}{\partial\gamma}}{(\Gamma(N\gamma))^2}}{\sum_{N=1}^{\infty}\binom{N+r-1}{N}q^N\frac{(\nu z_i)^{N\gamma}}{\Gamma(N\gamma)}} = 0 \qquad (5.4)$$

Limiting the summation of numerator for $N = 1$ and $N = 2$ that of denominator for $N = 1$, we get following equations after simplification.

$$p^2(A\sum_{i=1}^{n} {z_i}^{\gamma}) - p(A\sum_{i=1}^{n} {z_i}^{\gamma} + rn + n) + rn = 0 \qquad (5.5)$$

Where $A = \frac{r+1}{2}\frac{\Gamma(\gamma)}{\Gamma(2\gamma)}\nu^\gamma$ and r is arbitrary number 1, 2, 3, …

$$(B\gamma\nu^\gamma - \nu)\bar{z} + \gamma = 0 \qquad (5.6)$$

Where $B = \frac{\Gamma(\gamma-1)}{\Gamma(2\gamma-1)}\frac{r+1}{2}(1-p)$

$$log\nu + log\tilde{z} - \frac{\partial log\Gamma(\gamma)}{\partial\gamma} = 0 \qquad (5.7)$$

6. Conclusion

Not only we have introduced the Compound Poisson Frailty model and Compound Negative Binomial Frailty Model but we have tried to get maximum likelihood estimators. We get intrinsic equations.

7. References

1. Aalen OO. Heterogeneity in survival analysis. Statistics in Medicine 1988; 7:1121-1137.
2. Aalen OO. Modelling heterogeneity in survival analysis by the compound Poisson distribution. Annals of Applied Probability 1992; 4:951-972
3. Aalen OO, Tretli S. Analyzing incidence of testis cancer by means of a frailty model. Cancer Causes and Control 1999; 10:285–292
4. Hougaard P. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. Biometrika, 1984; 71(1):75-83.

5. Hougaard P. Survival models for heterogeneous populations derived from stable distributions. Biometrika, 1986a; 73:387–396.
6. Hougaard P. A class of multivariate failure time distributions. Biometrika 1986b; 73:671–678.
7. Hougaard P, Myglegaard P, Borch-Johnsen K. Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. Biometrics 1994; 50:1178–1188.
8. Haukka J, Suvisaari J, Lonnqvist J. Increasing age does not decrease risk of schizophrenia up to age 40. Schizophrenia Research. 2003; 61:105–110.
9. Manton K, Stallard E, Vaupel J. Alternative models for heterogeneity of mortality risks among the aged. Journal of the American Statistical Association 1986; 81:635-644.
10. Moger TA, Aalen OO, Halvorsen TO, Storm HH, Tretli S. Frailty modelling of testicular cancer incidence using Scandinavian data. Biostatistics 2004a; 5:1-14.
11. Vaupel JW, Yashin AI. Heterogeneity's ruses: Some surprising effects of selection on population dynamics. The American Statistician, 1985; 39:176–185.