Geometric Model Generalized For Cardiovascular Studies

Alfred A. Bartolucci[1], Ramalingam Shanmugam[2], Karan P.Singh[1]

1. Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Al. USA
2. Department of Mathematics, University of Colorado, Denver, CO. USA

Cardiovascular and related respiratory diseases are serious and leading causes of death. Given the gravity of the disease, this phenomenon has drawn the attention of the medical community to seek ways to improve the treatment techniques. Biostatisticians in turn have developed appropriate statistical methodologies to analyze data from cardiovascular patients. To better explain data on this phenomenon, the geometric model is generalized in truncated form. Statistical properties of this new model, expressions to estimate the model parameters and a statistic to test the "heart failure rate" are derived. The heart failure rate is seen as analogous to the statistical hazard rate. The new model is called the intervened geometric distribution (IGD). The model itself is generalized to the intervened geometric distribution of the nth order (IGDN). Both models can be used to test the effectiveness of any intervention. An illustrative example is provided.

1.0 Background and Motivation

Let Y be a random number of heart attacks suffered by an individual before coming to a hospital for cardiovascular treatment. Realize that Y=0 is not observable. The observable sample space for Y is the set of positive integers {1,2,.....}. Hence , a natural and reasonable choice for describing the uncertainty of Y is a zero truncated geometric (ZTG) distribution,

$$P[Y=y|\theta]=(1-\theta)\theta^{y-1} \quad ; \quad y=1,2,..... \qquad (1)$$

where $\theta\epsilon(0,1)$ denotes the chance of having heart failure in a unit time interval. Notice that the expectation of Y is $(1-\theta)^{-1}$ or on the average the number of heart failures a cardiovascular patient entering a hospital might have is $(1-\theta)^{-1}$.

Depending on the complexity of the illness, a cardiovascular patient might undergo either surgical or a medical treatment. Whatever the treatment may be, it will have an effect of changing the incidence parameter θ into a new setting $\rho\theta$ for a patient during the post treatment period. That is the average number of heart failures to be suffered by a patient during a period after the treatment is $E(Z)=\rho\theta$, a function of $\theta, \rho \in [0, \infty)$ where Z denotes the number of heart failures to be suffered by a patient during the post treatment period, and ρ is interpreted as the "treatment" effect. Z follows a geometric distribution:

$$Pr[Z=z|\rho,\theta)] =(1-\rho\theta)(\rho\theta)^z \; ; \; z=0,1,2,... ; \; \rho\theta<1 \qquad (2)$$

Notice that unlike Y, zero is a possibility for the random variable, Z. The zero value for ρ is indicative of successful treatment in the sense that the expected number of heart failures after treatment is zero. Whereas $\rho=1$ is indicative of the status quo after treatment. In situations in which the heart failure incidence is more after treatment ρ will be greater than one.

2.0 The Intervened Geometric Distribution

Suppose that an observational apparatus has kept a recordf only X=Y+Z, namely the total number of heart failures suffered by a patient altogether in one's entire lifetime until death which includes both pre and post treatment periods. Assume that the R.V's Y and Z are independent. The probability function (PF) of X is then

$$P[X=x|\rho,\theta]=\sum_{i=1}^{x} P[Y=i]P[Z=x-i|Y=i] \qquad (3a)$$

which yields

$$P[X=x|\rho\neq 1,\theta]=(1-\theta)(1-\rho\theta)\bullet \left|(\rho^{x}-1)/(\rho-1)\right|\theta^{x-1} \quad , \qquad (3b)$$

and

$P[X=x|\rho=1,\theta]=x(1-\theta)^2\theta^{x-1}$,

where x=1,2,3...., $\theta\in(0,1)$, and $\rho\in[0,\infty)$. This is the "intervened geometric distribution" (IGD). The mean, μ_x , and variance, σ^2_x, of the IGD are easily obtained and they are:

$$\mu_x=E(X)=(1-\rho\theta^2)/(1-\theta)(1-\rho\theta), \qquad (4)$$

and

$$\sigma^2_x=var(X)=(\theta/(1-\theta^2)[1+\rho((1-\theta)/(1-\rho\theta))^2] . \qquad (5)$$

The median, X_{me}, of IGD in (3) could be obtained from its survivor function, $S_x(a,\rho,\theta)=P[X>a|\rho,\theta]$, which is $1-F(a|\rho,\theta)=P[X\leq a|\rho,\theta]$. That is

$$F(a|\rho\neq 1,\theta)=(1-\theta)(1-\rho\theta)/(\theta(\rho-1)\sum_{x=1}^{a}(\rho^{x}-1)\theta^{x}. \qquad (6)$$

The survivor function of IGD is:

$$S_x(a|\rho\neq 1,\theta)=\theta^{a}[((\rho^{a+1}(1-\theta)-(1-\rho\theta))/(\rho-1)]. \qquad (7)$$

The case in which $\rho=1$could be discussed using the L'Hospital rule on these results since

$$\lim_{\rho\to 1} P[X=x|\rho\neq 1,\theta]=P[X=x|\rho=1,\theta] \text{ in}$$

(3). In other words, the survival function in the case of $\rho=1$ is the limit as $\rho\to 1$ in (7) or

$$S_x(a|\rho=1,\theta)=\theta^{a}[1+(1-\theta)a]. \qquad (8)$$

To find the median , x_{me} , we equate the survival function in (7) or (8) to ½ depending on whether the treatment effect, ρ, is or is not equal to one. That is we rewrite the median finding equation as

$2(1-\theta)\theta^{X}me[\rho^{X}me^{+1}-(1-\rho\theta)/(1-\theta)]=(\rho-1)$,

when $\rho>1$, and recognizing that $1-\rho\theta<1-\theta$, we find an approximate but simple equation that

$2\theta^{X}me(1-\theta)\rho^{X}me^{+1}=(\rho-1)$,

which we solve to obtain the median,

$$X_{me}=\ln\{(\rho-1)/2\rho(1-\theta)\}/\{\ln\rho+\ln\theta\}, \text{ when } \rho>1. \qquad (9)$$

Similarly, when $\rho<1$ then $1-\theta<1-\rho\theta$ and an approximate solution for the median in this case is

$$X_{me}\simeq\left[\ln\{(1-\rho)/2(1-\theta)(1-\rho\theta)\}/\ln(\theta)\right], \text{ when } \rho<1. \qquad (10)$$

In the case that $\rho=1$ the following holds,

$$X_{me}\ln\theta+\ln X_{me}\simeq -\ln 2. \qquad (11)$$

The mode , X_{mo}, of the IGD in (3) could be obtained by solving the inequality that $Pr[X=x_{mo}-1]<Pr[X=x_{mo}]>Pr[X=x_{mo}+1]$. It yields after simplifications that:

$$X_{mo}\simeq \begin{cases} \ln[(1-\theta)/(1-\rho\theta)]/\ln\rho, & \text{when } \rho>1, \\ \ln[\rho(1-\theta)/(1-\rho)]/\ln\rho & \text{when } \rho<1, \\ \text{gli of } [1/(1-\theta)] & \text{when } \rho=1, \end{cases} \qquad (12)$$

where gli is greatest largest integer.

3.0 The Hazard Function

The concept of hazard rate (or sometimes called failure rate) function at time t is useful in survival analysis. In the case of a discrete distribution, it is defined as,

$h_x(t)=Pr[X=t]/Pr[X\geq t+1]$.

In cardiovascular studies, it is called the heart failure rate which for a patient during a period before treatment is

$$h_y(t)=Pr[Y=t]/S_y(t)=\frac{1}{\theta}(\frac{1}{\theta}-1) \qquad (13)$$

since

$$S_y(t)=1-(1-\theta)\sum_{y=1}^{t}\theta^{y-1}=\theta^{t+1}$$

and the heart failure rate is

$$h_z(t)=\Pr[Z=t]/S_z(t)=[(1/\rho\theta)-1] \qquad (14)$$

during the post treatment period since

$$S_z(t)=1-(1-\rho\theta)\sum_{z=0}^{t}(\rho\theta)^z=(\rho\theta)^{t+1}\ .$$

By comparing (13) and (14) we notice that the heart failure rate for a patient after the treatment is less than the heart failure rate before the treatment if

$$\rho>\theta(1+\theta)/(1+\theta^3). \qquad (15)$$

When we consider the heart failure rate at a time, t, for a patient who has been treated, it is either

$$h_x(t|\rho\neq 1)=\frac{\Pr[X=t]}{S_x(t)}=\frac{(\rho^t-1)(1-\rho\theta)}{\theta^2[\rho^{t+2}-(1-\rho\theta)/(1-\theta)} \qquad (16)$$

when there is a change in heart failure rate due to treatment, or

$$h_x(t|\rho=1)=\frac{t(\frac{1}{\theta}-1)^2}{[1+(t+1)(1-\theta)} \qquad (17)$$

when the status quo of the heart failure rate prevails even after the treatment. A comparison of (13) and (17) reveals that

$$h_x(t|\rho=1)<\theta^2[h_y(t)]^2,$$

since

$\frac{t}{[1+(t+1)(1-\theta)}<1$. Furthermore, because

$\theta\epsilon(0,1)$, we may state that $h_x(t|\rho=1)<[h_y(t)]^2$

implying that the heart failure rate until death for a patient who has undergone a treatment is less than the square of the heart failure rate before the treatment, when the effect has been of the status quo, i.e. $\rho=1$.

In general , when $\rho\neq 1$, it is interesting to note the relationship between $h_x(t)$, the overall heart failure rate of a patient who has undergone treatment and that of $h_y(t)$ and $h_z(t)$, the heart failure rate before and after treatment respectively. That is

$$h_x(t|\rho\neq 1)=\frac{(\rho^t-1)}{[\theta^2\rho^{(t+2)}(1+\frac{1}{h_x(t)})-\frac{1}{h_y(t)}]}. \qquad (18)$$

Notice the the treatment is most effective (i.e. when $\rho=0$), then (18) results in $h_x(t)=h_y(t)$ implying that the heart failure rates or hazards are indentical.

In a situation in which there is no change, i.e. $\rho=1$, even after the treatment, $h_z(t)$ is less than $h_y(t)$ since

$h_z(t)=\frac{1}{\theta}-1=\theta h_y(t)$ and in which $\theta\epsilon(0,1)$.

For computing the probability for specified values of the parameters, ρ and θ, the following recurrence relation of the IGD in (3) is noted,

$$(\rho^x-1)\Pr[X=x+1|\rho\neq 1,\theta]=\theta[\rho^{x+1}-1]\Pr[X=x|\rho\neq 1,\theta]. \qquad (19)$$

For $\rho=1$ we use L'Hospital's rule in (19) and obtain

$$x\Pr[X=x+1|\rho=1,\theta]=(x+1)\theta\Pr[X=x|\rho=1,\theta]$$

which provides an interpretation for θ.

4.0 The Intervened Geometric Distribution of the n^{th} Order

We now investigate the distribution of the sum

$T = \sum_{i=1}^{n} X_i$ of a random sample, $X_1, X_2, \ldots\ldots X_n$ of size $n \geq 1$ from an IGD in (3). Using the moment generating function,

$$\phi_x(u) = E(e^{ux}) = \sum_{x=1}^{\infty} e^{ux} \Pr[X = x]$$

$$= \frac{(1-\theta)(1-\rho\theta)}{(1-\theta e^u)(1-\rho\theta e^u)}$$ and changing e^u to w,

we obtain the probability generating function (PGF), $E(W^X)$ of X , which upon raising to the nth power, yields the PGF of T. Expanding the PGF of T in powers of w, we obtain the PF of T, That is

$$\Pr[T = t | n, \rho \neq 1, \theta] = \left[\frac{(1-\theta)(1-\rho\theta)}{\theta(\rho-1)}\right]^n H(t,n,\rho)\theta^t \quad (20)$$

where $\theta \in (0,1)$, $\rho \in [0,\infty)$, n=1,2,....., t=n,n+1,....., and $H(t,n,\rho)$ is a generalized geometric number. Some properties of this new number are:

$$H(t,n,\rho) = (-1)^n \binom{t-1}{n-1},$$

$$H(t,0,\rho) = \begin{cases} 1 & \text{if } t=0 \\ 0 & \text{if } t \neq 0, \end{cases} \quad (21)$$

$$H(t,1,\rho) = \rho^t - 1.$$

The number $H(t,n,\rho)$ quickly becomes very large even for moderate values t and n, and hence an approximation is worthwhile. Thus after some derivation,

$$H(t,n,\rho) \approx (\rho-1)^n \binom{t-1}{n-1} \rho^{t-n}, \quad (22)$$

The PF in (20) is called then intervened geometric distribution of the n^{th} order (IGDN).

5.0 Parameter Estimation

Consider a random sample $X_1, X_2, \ldots\ldots X_n$ of size $n \geq 1$ from a population with IGD in (3), then the log- likelihood function $\psi(x_1, x_2, \ldots\ldots x_n)$ of the sample is

$$\sum_{i=1}^{n} \psi(x_i | \rho \neq 1, \theta) = \sum_{i=1}^{n} \{\ln|\rho^{x_i} - 1| + n\bar{x}\ln\theta - n[\ln(1-\theta) - \ln\theta + \ln(1-\rho\theta) - \ln|1-\rho|]\}, \quad (23)$$

where $\bar{x} = \sum_{i=1}^{n} x_i / n$, the sample mean.

Differentiating separately with respect to θ and ρ in (23) and equating them to zero, we obtain the maximum likelihood estimates (MLE) $\hat{\theta}$ and $\hat{\rho}$ of the model parameters, θ and ρ. .The estimating equation for the incidence parameter, θ, is then

$$\bar{x} = 1 + \left(\frac{\theta}{1-\theta}\right)\left[(1-\rho\theta) + \frac{\rho(1-\rho\theta^2)}{(1-\rho\theta)}\right]. \quad (24)$$

In (24) we incorporate (4) with $\bar{x}$ in place of E(X) to simplify (24) to an approximation which is

$$\hat{\theta} \cong \frac{(\bar{x}-1)}{(\rho+1)\bar{x}} \quad (25)$$

after ignoring the term $\rho\theta^2(\bar{x}+1)$ which is much less than one. The MLE in (25) is root -n consistent.

Similarly , we solve the estimating equation,

$$\sum_{i=1}^{n} x_i \frac{\rho^{x_i} - 1}{(\rho^{x_i} - 1)} = \frac{n(1-\theta)}{(\rho-1)(1-\rho\theta)} \quad (26)$$

for estimating the treatment effect, ρ. Using equation (26) we derive,

$$\hat{\rho} \cong \frac{\bar{x}}{x-(1-\theta)}, \text{ when } \rho>1, \qquad (27)$$

and

$$\hat{\rho} \cong \frac{[\bar{x}(1-\theta)-1]}{\theta}, \text{ when } \rho<1. \qquad (28)$$

The MLE's, $\hat{\theta}$ and $\hat{\rho}$ are asymptotically unbiased.

6.0 Testing Whether The Treatment Was Effective

A zero value for ρ is indicative of completely successful treatment, whereas ρ=1 is to be interpreted as the status quo in the heart failure rate. Of interest to any medical team is whether effective treatment took place or at least the risk of heart attack was not increased. That is to say that it is of interest to test H_0:ρ=1 versus H_1:ρ≠1 (hopefully ρ<1) based on a random sample $X_1, X_2, \ldots\ldots X_n$ of size n from an IGD as in (3). Since the incidence rate, θ, is unknown in a real life situation, the testing of ρ is not straight forward. For this purpose we employed the Neyman's c[α] test procedure, and also use it to find a confidence interval estimate for the treatment effect. See Neyman (1959). Using this procedure an asymptotic test statistic is

$$\chi^2_{1df} = \frac{(\rho+1)\sum_{i=1}^{n}(x_i-\bar{x})^2}{(n-1)(\bar{x}-1)} \qquad (29)$$

which is chi squaredistributed on 1 degree of freedom under a hypothesized value of ρ. That is for a large sample we will reject the null hypothesis H_0: ρ=1, in favor of H_1: ρ<1, whenever the test score ,

$\chi^2_{1df} = \frac{2\sum_{i=1}^{n}(x_i-\bar{x})^2}{(n-1)(\bar{x}-1)}$ is smaller than the standard chi square percentile for some preselected significance level α in the interval (0,1). In the case of small sample size, critical values of $\chi_{\rho=1}$ should be computed based on simulated results. Also the power of our test score could be computed under a clearly specified alternative hypothesis. For this purpose, suppose that H_1:ρ=ρ*=1-Δ, where Δ∈(0,1) is a desired impact of the treatment effect.. Then one can see that:

$$Power = \Pr[\chi_{\rho=1} > \chi_\alpha | H_1: \rho = \rho^* = 1-\Delta \in (0,1)$$

$$= \Pr\left[\chi_{\rho^*} > \chi_\alpha \left(\frac{1-\frac{\Delta\bar{x}}{x+1}}{1-\frac{\Delta}{2}} \right)\right] \qquad (30)$$

where χ_{ρ^*} is chi square distributed on one degree of freedom under H_1.

The 100(1-α)% confidence interval for the treatment effect , ρ, we proceed as follows. Note that

$\Pr[-Z_{\alpha/2} < Z_\rho < Z_{\alpha/2}] = 1-\alpha$ in which

$Z_\rho = \sqrt{\chi^2_{1df}}$ in (29) and $Z_{\alpha/2}$ is the $100(\alpha/2)^{th}$ percentile from the standard normal distribution. We note that $Z^2_{\alpha/2} = \chi^2_{\alpha/2,1df}$, the chi squared percentile with one degree of freedom (df). Also we write $\frac{(\rho\bar{x}+1)}{\sqrt{(\rho+1)\bar{x}}}$ as $\rho\sqrt{\frac{\bar{x}}{(\rho+1)}} + \frac{1}{\sqrt{(\rho+1)\bar{x}}}$ and recognize that the second term is negligible since $\bar{x} \geq 1$. The coefficient $\frac{\rho}{\sqrt{\rho+1}}$ in the first term could be approximated to $\sqrt{\rho+1}$ after writing it as $\sqrt{\rho+1} - \frac{1}{\sqrt{\rho+1}}$ and recognizing that the second term is negligible. Hence we write an approximate probability ,

$$\Pr\left[-Z_{\alpha/2} < \frac{\sqrt{(\rho+1)\bar{x}}[\sum(x_i-\bar{x})^2]}{\sqrt{(n-1)(\bar{x}-1)}} < Z_{\alpha/2}\right]$$

$\cong 1-\alpha.$

By squaring both sides of this expression , rearranging terms , and defining

$$B = \frac{2\bar{x}(n-1)(\bar{x}-1)}{[\sum (x_i - \bar{x})]^2}$$, we have,

$$\Pr\left[B\chi^2_{\alpha/2,1df} - 1 < \rho < B\chi^2_{1-\alpha/2,1df} - 1\right] \cong 1-\alpha. \quad (31)$$

7.0 Numerical Example

We have data from three different time periods at a hospital treating cardiology patients. Table 1 summarizes these results from the data gathered. We call the three data sets "studies". The data sets are fairly small and thus the large sample approximations may not be entirely appropriate. However, they do demonstrate the results rather nicely. Table 2 follows with the statistical test results and the confidence limits. One can see in the first two studies that based on the chi square statistics one would reject the null at least at the 0.10 level in that the treatment may have been effective. The last result demonstrates that the treatment was in fact not effective. One can also see from the estimates of θ that the prior heart failure rate was lower initially in the first two studies. The confidence intervals for ρ are rather wide and do encompass the observed parameter values rather easily. The MLE's demonstrate the overall true outcomes in each case.

Study	N	$\bar{X}$	$\hat{\theta}$	$\hat{\rho}$
1	15	1.20	0.106	0.573
2	20	1.20	0.031	0.553
3	20	1.80	0.154	1.89

Table 1: Parameter Estimates.

Study	χ^2_{1df}	Lower Limit	Upper Limit
1	3.77	0	5.38
2	3.00	0	5.29
3	4.07	0	7.87

Table 2: Chi Square and Confidence Limits for ρ.

8.0 Conclusions

Overall the methodology does demonstrate the value of the model in that now one does have an analytic tool in the IGD to actually measure the post treatment effectiveness for heart attack victims. The appropriate approximations do make the equations tractable. The danger in use of some approximations is that one can miss results that may really be quite sensitive to the change in the status pre to post treatment. We have been fortunate to be able to demonstrate this procedure. Larger data sets will also be needed from other Medical Centers to further validate this procedure.

9.0 References

Kannal,W.B., Wilson,P.,Blaire,S.N., Epidemiological assessment of the role of physical activity and fitness in the development of cardiovascular disease, *American Heart Association Journal*, 109, 876-885, 1985.

Kendall, M.G., Stuart, A, *The Advanced Theory of Statistics. Volumes 1 and 2*,Griffin Publication. London, 1990.

Neyman,J, Optimal asymptotic tests of composite hypotheses in probability and statistics, *The Harold Cramer Volume*. John Wiley Press, New York, 1959.

Peters, T.J and Golding, J, Prediction of sudden death syndrome: an independent evaluation of four scoring methods. *Statistics in Medicine*, 5, 113-126, 1986.