International Journal of Research in Medical Sciences
Shanmugam R. Int J Res Med Sci. 2015 Sep;3(9):2181-2189
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012

DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20150599

Research Article

Learning from an early start but late end epidemics via an incidence rate restricted bivariate distribution and data analysis

Ramalingam Shanmugam*

Department of Health Administration, Texas State University, San Marcos, TX 78666, USA

Received: 31 July 2015
Revised: 01 August 2015
Accepted: 11 August 2015

***Correspondence:**
Dr. Ramalingam Shanmugam,
E-mail: rs25@txstate.edu

ABSTRACT

Background: An ideal expectation of public health administrators or field medical workers is to have a late start and quick ending of any epidemic. Instead, when an epidemic starts early but ends late, it is where much can be learned from the incidences. A case in point for discussion in this article is the pattern of 2009 H1N1 epidemic.
Methods: With a parameter to portray an existing health environment as a deterrent for an epidemic like H1N1 to outbreak in any location at a week, a bivariate distribution is created and is used to analyze the data for a learning so that it helps to prevent a too long prevailing future epidemic. This new distribution is named Incidence Rate Restricted Bivariate Distribution (IRRBGD). Statistical properties of IRRBGD are derived and illustrated using 2009 H1N1 incidences in all five continental regions (Africa, Asia, Europe, Americas, and Oceanic) across on earth.
Results: The Asian continent, compared to other four continental regions, had most vulnerability for H1N1 incidences. The odds for no H1N1 to occur is lowest only in Oceanic among the four continental regions, namely Africa, Europe, Americas, and Oceanic. Since the beginning of the year 2009 with 52 weeks, the week number, Y in which the H1N1 appeared first and the number, X of weeks the H1N1 continued on in a region are consistently highly correlated in all five continental regions.
Conclusions: From the data analyses of 2009 H1N1 incidences, no continental region is risk free with respect another round of H1N1 epidemic in future. The medical community and public healthcare administrators ought to identify the common and region specific unique deterrents of the epidemic like H1N1. The impact of such deterrents to H1N1 is captured in our model and analysis. By increasing the deterrent level, the outbreak of an epidemic like H1N1 could be delayed, according to our model and data information.

Keywords: Conditional discrete distribution, Two parameter geometric distribution, Survival function, Correlation, Regression

INTRODUCTION

The World Health Organization (WHO) worried about the recurrence pf the 2009 outbreak of H1N1 pandemic worldwide as it consumed then many human lives. What is H1N1? The disease H1N1 is a swine Flu (as it expresses H or N antigens). The H1N1 case is confirmed only by a lab test and not by its symptoms: fever, sore throat, nasal congestion, cough, respiratory problems, or body aches. Smith et al.[1] narrates the origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. The name "swine" was selected because people caught it first in direct contact with the pigs. Why do we have to understand H1N1 pandemic thoroughly? The health/medical professions could prevent such pandemic in future. On June 11, 2009, WHO declared that H1N1 was the first global pandemic since the 1968 Hong Kong flu. Saito et al.,[2] Jain et al.,[3]

Smith et al.,[1] and Vijaykrishna et al.[4] provide details about the 2009 H1N1 pandemic.

The H1N1 virus is destroyed only by heat at a high-level 167-212°F (equivalently, 75-100°C). On October 25, 2009, the U.S. President Barack Obama declared that H1N1 was a national emergency. In May 2009, the Chinese government confined 21 U.S. students and 3 teachers to their hotel rooms. Australia ordered a cruise ship with 2000 passengers to stay at sea because of a swine flu. Japan quarantined 47 airline passengers in a hotel for a week in mid-May. In mid-June 2009, India ordered pre-screening the "outbound" passengers from the countries thought to have a high rate of infection. Pregnant women have a higher risk to be a H1N1 case. About 18,138 deaths occurred worldwide before the start of year 2010. Only on August 10, 2010, the WHO declared that the H1N1 pandemic is over. Hence, this article gives an importance to this pandemic and analyzes the 2009 data (http://www.cdc.gov/h1n1flu/qa.htm).

A closer look at the 2009 H1N1 incidences makes one wonder why a late start of H1N1 in a nation ends early. What triggers the H1N1 to start late? Their relationship results in a negative correlation in all five continents: Africa, Asia, Europe, Americas, and Oceanic. Is it because there exist deterrents to the epidemic in the nations of a continent or knowledge of medical expediencies based on learning from other countries past performance to extinguish the pandemic H1N1?

How should a data analysis proceed to answer the above question? Definitely, an appropriate underlying model for the data needs to be first identified. What is a model? Model is an abstraction of the reality. A search of the literature reveals that there is no appropriate bivariate distribution currently available in the literature to serve as an underlying model for the 2009 H1N1 data.

Hence, this article first develops a new bivariate discrete distribution to suit the H1N1 data and names it incidence Rate Restricted Bivariate Geometric Distribution (IRRBGD). The statistical properties IRRBGD and a methodology based on those properties to analyze data and answer the above stated question are done in this article.

An illustration of all derived new results of this article is made using the 2009 H1N1 incidences in five continents: Africa, Asia, Europe, Americas, and Oceanic. The similarities and the differences among the five continents are compared and commented in the end. A few thoughts are pointed out for future research directions in the end.

DERIVATION OF INCIDENCE RATE RESTRICTED BIVARIATE GEOMETRIC DISTRIBUTION

First, we need to derive a new discrete bivariate probability distribution with reasons. To be specific, let Y be a random variable denoting the *number of elapsed weeks* since the beginning of the clock when the pandomic H1N1 first appeared in a nation with a chance $0<\theta<1$. The possibilities for Y are integers $\square^{+}=\{0,1,2,3,....\}$. A natural candidate to be the underlying *probability mass function* (PMF) of Y is the *geometric distribution* (GD) and it is

$$\Pr[Y=y|\theta]=(1-\theta)\theta^{y}; y=0,1,2,3,...,,; 0<\theta<1 \quad (1)$$

The *odds* for no H1N1 to occur in a week is the ratio

$$\text{Odds}_{\theta}=\frac{\Pr[Y=0]}{1-\Pr[Y=0]}=\frac{1}{\theta}-1 \quad (2)$$

The Odds_{θ} increases if the incidence rate θ decreases. The *expected number,* $E[Y=y|\theta]$ of elapsed weeks after which the H1N1 occurs and the *volatility* (which is recognized as *variance,* $\text{Var}[Y=y|\theta]$ in mathematical statistics) are respectively

$$\mu_{\theta}=E[Y=y|\theta]=\sum_{y=1}^{\infty}y(1-\theta)\theta^{y}=\frac{\theta}{(1-\theta)}=\frac{1}{\text{Odds}_{\theta}}, \quad (3)$$

and

$$\sigma_{\theta}^{2}=\text{Var}[Y=y|\theta]=\sum_{y=1}^{\infty}[y-\frac{(1-\theta)}{\theta}]^{2}(1-\theta)\theta^{y}=\mu_{\theta}(\mu_{\theta}+1) \quad (4)$$

The expected number of weeks happened to be the inverse of the odds of H1N1 outbreak. The mean μ_{θ} number of elapsed weeks for H1N1 to occur first increases when the incidence rate θ increases. The volatility, σ_{θ}^{2} increases when the mean, μ_{θ} increases as their relationship (6) is parabolic. The most likely (that is, *mode*) is $\mu_{\text{mode}}=0$. The GD (1) is skewed and the amount of *skewness* is $S_{k}=\sqrt{\theta}+\frac{1}{\sqrt{\theta}}$. An *excessive kurtosis* (which is tail flatness in the frequency trend) exists and is $K_{u}=4+\theta+\frac{1}{\theta}$. The *entropy* of GD (1) is $\Im_{\theta}=[-\frac{\theta}{(1-\theta)}\ln\theta-\ln(1-\theta)]$. The probability that the *number,* Y *of elapsed weeks* since the beginning of clock for H1N1 to occur exceeds the $(w+1)^{\text{th}}$ week is the survival function and it is

$$S_{w}=\Pr[Y\geq(w+1)]=(1-\theta)\sum_{y=(w+1)}^{\infty}\theta^{y}=\theta^{(w+1)} \quad (5)$$

The survival probability (5) becomes slim when "w" increases. In other words, the chance for H1N1 to occur is more in an earlier week than in the later weeks. Nevertheless, the 2009 H1N1 data contradicts it and it necessitates a need to modify the GD (1) as it is done below.

Chiolero et al.[5] has advocated the versatility and importance of generalizations. If a pandemic like H1N1 has a late start in a location, it must have been due to a *deterrent level:* $0<\theta<\phi\le 1$. The probability for a H1N1 incidence to occur in a week under the framework of a *deterrent level* is dampened to a low level, $\theta(1-\theta)^{(\frac{1}{\phi}-1)}$ where it is clear that $\theta(1-\theta)^{(\frac{1}{\phi}-1)}<\theta$. In other words, the parameter ϕ portrays the *deterrent level* at the location, which delays the H1N1 epidemic. By fusing in the *deterrent* level into the GD (1), the probability model becomes more versatile and extra informative. An extension of the GD (1) to consider is

$$\Pr[Y=y|\theta,\phi]=[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}][\theta(1-\theta)^{(\frac{1}{\phi}-1)}]^{y};$$
$$0<\theta<\phi\le 1; y=0,1,2,3,...,; \quad . \quad (6)$$

Is the expression (6) bona-fide PMF? First, note that clearly, $\theta(1-\theta)^{(\frac{1}{\phi}-1)}<1$ because $0<\theta<1$ and $0<\phi\le 1$. Consequently, $0<\theta(1-\theta)^{(\frac{1}{\phi}-1)}<1$. Furthermore,

$$\sum_{y=0}^{\infty}\Pr[Y=y|\theta,\phi]=[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}]\sum_{y=0}^{\infty}[\theta(1-\theta)^{(\frac{1}{\phi}-1)}]^{y}$$
$$=[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}][\frac{1}{1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}}]=1 \quad .$$

Hence, the expression (6) is indeed a bona-fide PMF. Being new to the literature, the PMF (6) is named *incidence rate restricted geometric distribution (IRRGD)* in this article.

When the *deterrent level* is at its maximum, (that is, $\phi=1$), the IRRGD (6) reduces to the GD (1) as a special case. This is truly ideal situation but impractically fictitious, because no nation really is at a complete deterrent. Hence, using the GD (1) amounts to making a preponderous assumption that the nation offers a *full deterrent* with respect to H1N1 occurrence. The value $\phi=1$ is recognized as the *target baseline level.* Otherwise, the *odds* for no H1N1 to occur in a week is the ratio

$$\text{Odds}_{\theta,\phi,Y=0}=\frac{\Pr[Y=0|\phi]}{1-\Pr[Y=0|\phi]}$$
$$=\frac{1}{\theta(1-\theta)^{(\frac{1}{\phi}-1)}}-1 \quad . \quad (7)$$

The $\text{Odds}_{\theta,\phi}$ (7) for no H1N1 to occur is $\frac{1}{(1-\theta)^{(\frac{1}{\phi}-1)}}$ percentage more than its counterpart Odds_{θ} (2) with a complete deterrent and it reflects the impact of the existing *deterrents* in the nation. The mean and variance of IRRGD (6) are derived and they are respectively

$$\mu_{\theta,\phi}=E[Y=y|\theta,\phi]=\frac{\theta(1-\theta)^{(\frac{1}{\phi}-1)}}{[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}]}$$
$$=[\frac{1}{(1+\mu_{\theta})^{\frac{1}{\phi}}-\mu_{\theta}}]\mu_{\theta} \quad (8)$$

and

$$\sigma^2_{\theta,\phi}=\text{Var}[Y=y|\theta,\phi]=\frac{\theta(1-\theta)^{(\frac{1}{\phi}-1)}}{[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}]^2}$$
$$=\mu_{\theta,\phi}(\mu_{\theta,\phi}+1)=\frac{\mu_{\theta}(1+\mu_{\theta})^{\frac{1}{\phi}}}{[(1+\mu_{\theta})^{\frac{1}{\phi}}-\mu_{\theta}]^2}$$
$$=[\frac{(1+\mu_{\theta})^{(\frac{1}{\phi}-1)}}{\{(1+\mu_{\theta})^{\frac{1}{\phi}-}-\mu_{\theta}\}^2}]\sigma^2_{\theta} \quad . \quad (9)$$

The mean $\mu_{\theta,\phi}$ (8) changes by an amount $[\frac{1}{(1+\mu_{\theta})^{\frac{1}{\phi}}-\mu_{\theta}}]$ from its counterpart mean μ_{θ} (3) due to the deterrent level, ϕ. When $\phi=1$, there is no change in the mean.

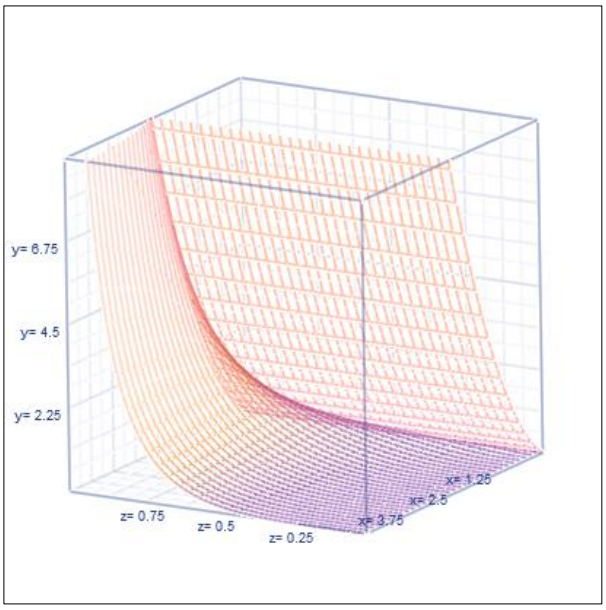

Figure 1: Volatility bends and twists due to the preventive protection.

Likewise, the change in the volatility $\sigma^2_{\theta,\phi}$ (9) from its counterpart σ^2_{θ} (4) is $\frac{(1+\mu_\theta)^{(\frac{1}{\phi}-1)}}{\{(1+\mu_\theta)^{\frac{1}{\phi}-}-\mu_\theta\}^2}$ due to the deterrent level ϕ against H1N1. The volatility, $\sigma^2_{\theta,\phi}$ of IRRGD (6) increases when the mean, $\mu_{\theta,\phi}$ increases as their relationship is $\sigma^2_{\theta,\phi}=\mu_{\theta,\phi}(\mu_{\theta,\phi}+1)$ like the volatility, σ^2_{θ} of GD (1) increases when the mean, μ_θ increases as they have a similar relationship $\sigma^2_{\theta}=\mu_\theta(\mu_\theta+1)$.

By letting the volatility in the $y-axis$, the mean μ_θ in the $x-axis$ and the deterrent level, ϕ in the $z-axis$, the volatility $\sigma^2_{\theta,\phi}$ (9) and the volatility σ^2_{θ} (4) are compared in Figure 1. An interpretation of the configurations (the bent and twisted is for $\phi\neq 1$ and the light-shaded is for $\phi=1$) is the following. Notice that when the volatility is more, the deterrent level, ϕ is more. The volatility steadily decreases in the absence of the deterrent but swiftly decreases in the presence of deterrent as the mean number of weeks for H1N1 to occur increases.

The *mode* is $\mu_{mode}=0$. It is clear that the IRRGD (6) is skewed and the amount of *skewness* is $S_k=\sqrt{\theta(1-\theta)^{(\frac{1}{\phi}-1)}}+\frac{1}{\sqrt{\theta(1-\theta)^{(\frac{1}{\phi}-1)}}}$. The *excessive kurtosis* is $K_u=4+\theta(1-\theta)^{(\frac{1}{\phi}-1)}+\frac{1}{\theta(1-\theta)^{(\frac{1}{\phi}-1)}}$. The *entropy* of the IRRGD (6) is

$$\Im_\theta=[-\frac{\theta(1-\theta)^{(\frac{1}{\phi}-1)}}{[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}]}\{\ln\theta+(\frac{1}{\phi}-1)\ln(1-\theta)\}$$
$$-\ln\{1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}\}]\quad .$$

The probability that the *number,* Y *of elapsed weeks* since the beginning of clock will exceed the $(w+1)^{th}$ week under the existing *deterrent level,* ϕ is the survival function and it is

$$S_{w,\phi}=\Pr[Y\geq(w+1)]$$
$$=[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}]\sum_{y=(w+1)}^{\infty}[\theta(1-\theta)^{(\frac{1}{\phi}-1)}]^y$$
$$=[\theta(1-\theta)^{(\frac{1}{\phi}-1)}]^{(w+1)}$$
$$=(1-\theta)^{(\frac{1}{\phi}-1)(w+1)}S_w \quad . \qquad (10)$$

It is clear from the *survival probability* $S_{w,\phi}$ in (10) that the H1N1 incidence is delayed to occur by an amount $[\theta(1-\theta)^{(\frac{1}{\phi}-1)}]^{(w+1)}$ due to the existing deterrent *level* ϕ.

We now follow a similar line of thinking as in Shanmugam[6], which was used to address the delayed recording of HIV/AIDS data in the reporting system. The probability shortfall for H1N1's delayed occurrence is $\frac{(S_w-S_{w,\phi})}{S_w}=[1-(1-\theta)^{(\frac{1}{\phi}-1)}]^{(w+1)}$. In other words, the shortfall in probability for the H1N1 not to occur in the beginning week is $[1-(1-\theta)^{(\frac{1}{\phi}-1)}]$.

Given that the H1N1 has not occurred by the end of $(w+1)^{th}$ week, the probability that H1N1 will not occur in the next week is the Markov chain with an *attained memory level,* $\theta(1-\theta)^{(\frac{1}{\phi}-1)}$ under an existing deterrent level $\phi\neq 1$ in comparison to its counterpart *ideal memory level* θ under a full deterrent level $\phi=1$.

Now, we discuss how long the H1N1 continues once it has occurred. To be specific, let X be a *random number of weeks* the pandemic H1N1 continues on once it started on the y^{th} week. For X, we could consider a *conditional geometric distribution* (CGD)

$$\Pr[X = x|Y = y,\lambda] = [1-e^{-\lambda y}]e^{-\lambda yx}; 0 < \lambda < \infty;$$
$$x = 0,1,2,3,....,\infty; y = 0,1,2,3,...,\infty \quad . \quad (11)$$

The *odds* for the H1N1 to *stop (not continuing on)* is

$$\text{Odds}_{x|y,\lambda} = \frac{\Pr[X = 0|Y = y,\lambda]}{1-\Pr[X = 0|Y = y,\lambda]} = \frac{[1-e^{-\lambda y}]}{[e^{-\lambda y}]} \quad . \quad (12)$$

The $\text{Odds}_{x|y,\lambda}$ (12) for H1N1 to stop (not continuing on) increases as the number y increases. Let the impact of the *healthcare administrators'* efforts to stop H1N1 be a parameter $\lambda > 0$. When $\lambda > 0$ increases, the $\text{Odds}_{x|y,\lambda}$ (12) for H1N1 to stop increases. Hence, we recognize the parameter $\lambda > 0$ as the *healthcare administrators' efforts*. The importance of the *healthcare administrators' efforts* is confirmed by the mean and variance of the *CGD* (11). The mean and variance are respectively

$$\mu_{x|y,\lambda} = E[X = x|y,\lambda] = \frac{[1-e^{-\lambda y}]}{e^{-\lambda y}} = \frac{1}{\text{Odds}_{x|y,\lambda}} \quad (13)$$

and

$$\sigma^2_{x|y,\lambda} = \text{Var}[X = x|y,\lambda] = \frac{e^{-\lambda y}}{[1-e^{-\lambda y}]^2} = (1+\mu_{x|y,\lambda}) \quad . \quad (14)$$

The condtional mean, $\mu_{x|y,\lambda}$ in (13) points out that the expected number of weeks for the H1N1 to continue on is more when the $\text{Odds}_{x|y,\lambda}$ for the H1N1 to stop is lesser and vice versa. Also, the *conditional volatility*, $\sigma^2_{x|y,\lambda}$ in (14) increases when the conditional mean number, $\mu_{x|y,\lambda}$ of weeks the H1N1 to continue on is more (that is, $\mu_{x|y,\lambda} \to \infty$).

The *mode* of CGD is $\mu_{x|y,\lambda} = 0$. The CGD (11) is skewed and the amount of *conditional skewness* is $S_k = \sqrt{e^{-\lambda y}} + \frac{1}{\sqrt{e^{-\lambda y}}}$. The *excessive kurtosis* of CGD (11) is $K_u = 4 + e^{-\lambda y} + e^{\lambda y}$.

The *entropy* is $\Im_\theta = \frac{1}{(e^{\lambda y}-1)}\{\lambda y + \ln\{1-e^{-\lambda y}\}$. Entropy is the statistical information. The entropy increases when either $\lambda \to \infty$, $y \to \infty$, or both.

The conditional probability for the H1N1 to continue on beyond the $(w+1)^{th}$ week without stopping once it started in the y^{th} week is the survival function and it is

$$S_{x|y}(w+1) = \Pr[X \geq (w+1)|y,\lambda] = e^{-\lambda y(w+1)} \quad . \quad (15)$$

In other words, we ask that given that the H1N1 has not stopped by the end of the $(w+1)^{th}$ week, what is the probability that H1N1 would not stop in the next week also. It is $e^{-\lambda y}$.

We now proceed to utilize the joint PMF of the random variables Y and X. From the marginal PMF (6) and the conditional PMF (11), we write the joint PMF and it is

$$\Pr[X = x, Y = y|\theta,\phi,\lambda,\tau] = \Pr[Y = y|\theta,\phi]\Pr[X = x|Y = y,\lambda]$$
$$= [1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}][\theta(1-\theta)^{(\frac{1}{\phi}-1)}e^{-\lambda x}]^y[1-e^{-\lambda y})];$$
$$x = 0,1,2,3,....,\infty; y = 0,1,2,3,...,\infty;$$
$$0 < \theta < \phi \leq 1; 0 < \lambda < \infty; 0 \leq \tau < \infty. \quad (16)$$

The expression (16) is named an *incidence rate restricted bivariate geometric distribution* (IRRBGD) in this article.

The marginal PMFs, conditional PMFs, conditional expected values, conditional variances, correlation, and regression of the *random variables* Y and X are next derived. The marginal PMF of Y is already displayed in (6). Next, we obtain the *marginal PMF of* X.

$$\Pr[X = x|\theta,\phi,\lambda]$$
$$= \sum_{y=0}^{\infty}\Pr[X = x, Y = y|\theta,\phi,\lambda]$$
$$= \sum_{y=0}^{\infty}[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}][\theta(1-\theta)^{(\frac{1}{\phi}-1)}e^{-\lambda x}]^y[1-e^{-\lambda y})]$$
$$= (1-e^{-\lambda})e^{-\lambda x};$$
$$x = 0,1,2,3,....,\infty; 0 < \lambda < \infty. \quad (17)$$

The expression (17) is a bona fide *geometric distribution*. Hence, the bonafide conditional distribution of Y given $X = x$ is

$$\Pr[Y = y|X = x,\theta,\phi,\lambda] = \frac{\Pr[X = x, Y = y|\theta,\phi,\lambda]}{\Pr[X = x|\theta,\phi,\lambda]}$$
$$= [\frac{1}{\theta(1-\theta)^{(\frac{1}{\phi}-1)}e^{-\lambda x}} - 1][\frac{1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}e^{-\lambda(x+1)}}{1-e^{-\lambda}}]$$
$$[\theta(1-\theta)^{(\frac{1}{\phi}-1)}e^{-\lambda x}]^y[1-e^{-\lambda y}]$$
$$y = 0,1,2,3,...,\infty; x = 0,1,2,3,....,\infty;$$
$$0 < \theta < \phi \leq 1; 0 < \lambda < \infty. \quad (18)$$

The expression (18) is not seen anywhere in the literature, and hence, it is named *Shanmugam's conditional geometric distribution* (SCGD) in this article. The SCGD (18) is versatile enough to explain many real life chance mechanisms of earthquake incidences and modern astronomy theories, which will be explored separately in future articles.

Now, we obtain the marginal and conditional *expected values* and the *volatilities.* The expected values are:

$$E[Y = y|\theta,\phi] = \frac{\theta(1-\theta)^{(\frac{1}{\phi}-1)}}{[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}]}$$
$$\approx \theta(1-\theta)^{(\frac{1}{\phi}-1)}[1+\theta(1-\theta)^{(\frac{1}{\phi}-1)}] \quad , \qquad (19)$$

$$E[X = x|\lambda] = \frac{e^{-\lambda}}{[1-e^{-\lambda}]}$$
$$\approx e^{-\lambda}[1+e^{-\lambda}] \quad , \qquad (20)$$

$$E[X = x|Y = y,\lambda] = \frac{e^{-\lambda y}}{[1-e^{-\lambda y}]}$$
$$\approx e^{-\lambda y}[1+e^{-\lambda y}] \quad , \qquad (21)$$

$$E[Y = y|X = x,\theta,\phi,\lambda]$$
$$\approx (1-\lambda)\theta(1-\theta)^{(\frac{1}{\phi}-1)} - \lambda\theta(1-\theta)^{(\frac{1}{\phi}-1)}x \quad . \qquad (22)$$

$$E[XY] = E[XE[Y|X = x]]$$
$$\approx (1-\lambda)\theta(1-\theta)^{(\frac{1}{\phi}-1)} - \lambda\theta(1-\theta)^{(\frac{1}{\phi}-1)}\frac{e^{-\lambda}}{(1-e^{-\lambda})} \qquad (23)$$

The variances are (26)

$$Var[Y = y|\theta,\phi] = \frac{\theta(1-\theta)^{(\frac{1}{\phi}-1)}}{[1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}]^2}$$
$$\approx \theta(1-\theta)^{(\frac{1}{\phi}-1)}[1+\theta(1-\theta)^{(\frac{1}{\phi}-1)}]^2 \quad , \qquad (24)$$

$$Var[X = x|\lambda] = \frac{e^{-\lambda}}{[1-e^{-\lambda}]^2} \approx e^{-\lambda}[1+e^{-\lambda}]^2 \quad , \qquad (25)$$

$$Var[X = x|Y = y,\lambda] = \frac{e^{-\lambda y}}{[1-e^{-\lambda y}]^2}$$
$$\approx e^{-\lambda y}[1+e^{-\lambda y}]^2 \quad , \qquad (26)$$

$$Var[Y = y|X,\theta,\phi,\lambda] = E[Y(Y-1)|X,\theta,\phi,\lambda]$$
$$+E[Y|X,\theta,\phi,\lambda] - \{E[Y|X,\theta,\phi,\lambda]\}^2$$
$$\approx \lambda\theta(1-\theta)^{(\frac{1}{\phi}-1)}(x+1) \quad , \qquad (27)$$

$$Cov[X,Y] = E[XY] - E[X]E[Y]$$
$$\approx \theta(1-\theta)^{(\frac{1}{\phi}-1)}[1 - \frac{\lambda}{(1-e^{-\lambda})} - \frac{e^{-\lambda})}{(1-e^{-\lambda})\{1-\theta(1-\theta)^{(\frac{1}{\phi}-1)}\}}] \qquad (28)$$

Hence, the correlation coefficient between the *number, Y of weeks elapsed before the first appearance of H1N1* in a country and the *number X of weeks* the H1N1 continued on until it stopped is

$$Corr[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} \approx -\sqrt{e^{-\lambda}\theta(1-\theta)^{(\frac{1}{\phi}-1)}} \quad . \qquad (29.a)$$

An estimate of the correlation is

$$\hat{Corr}[X,Y] \approx -\sqrt{(\frac{\bar{x}}{1+\bar{x}})(\frac{\bar{y}}{1+\bar{y}})} \quad . \qquad (29.b)$$

Next, we estimate the parameters of the IRRBGD (16) with a random sample $(x_1,y_1),(x_2,y_2),....(x_n,y_n)$. Let $\bar{y}$, s_y^2, $\bar{x}$, and s_x^2 denote their mean and variances. Then, their estimators are:

$$\hat{\theta} \approx \max[(1-\frac{\bar{y}}{s_y^2})(-\ln[\frac{\bar{y}}{1+\bar{y}}]), 0.99] \qquad (30)$$

$$\hat{\phi} \approx \max(1-\frac{\bar{y}}{s_y^2}, 0.99) \qquad (31)$$

and

$$\hat{\lambda} \approx -\ln(\frac{\bar{x}}{1+\bar{x}}) \quad . \qquad (32)$$

Following Maher et al.[7]'s concepts and tools of regression to the mean, we make the following assertions. Given that the *number, Y* of elapsed weeks after which the H1N1 first appeared, of interest to the healthcare administrators is the projection, $\hat{X}_{projection}$ of the *number of weeks* the H1N1 might continue on until it stopped. This projection is feasible with a suitable regression and it is

$$\hat{X}_{projection} = \bar{x} - (\frac{\bar{x}}{1+\bar{x}})\frac{(y-\bar{y})}{(1+\bar{y})^2} \qquad (33)$$

using (19), (20), (24), (25), (29), (30), (31), and (32). All the above expressions are illustrated in the next section.

ILLUSTRATION WITH 2009 H1N1 EPIDEMICS

Not all nations in each continental region are homogeneous. See the scatterplots (in Figure 2 through Figure 6) to realize that there are outlier nations in every continent. In fact, the nations in Oceanic are too heterogeneous to have any regular pattern in H1N1 incidences. Hence, for the nations in the Oceanic continental region, the incidence rate restricted geometric distribution is somewhat unsuitable. In our discussion, the Oceanic continent consists of Australia, Fiji, New Caledonia, and New Zealand.

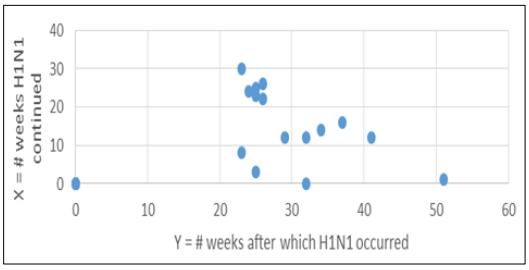

Figure 2: Regressive relation between X and Y in Africa ($\rho_{x,y}$ = -0.568 in Africa).

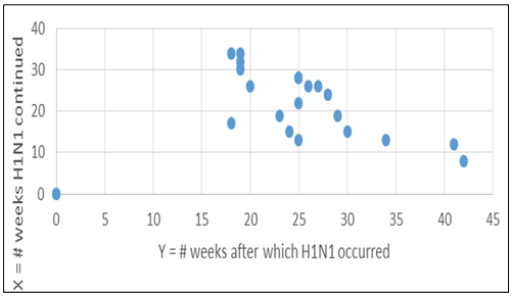

Figure 3: Regressive relation between X and Y in Asia ($\rho_{x,y}$ = -0.649 in Asia).

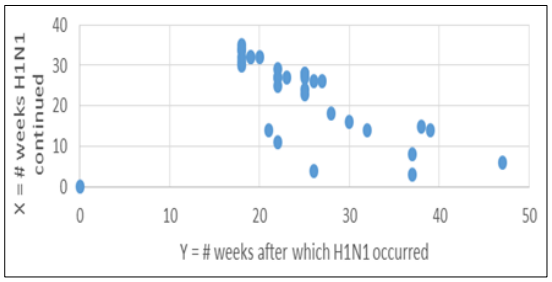

Figure 4: Regressive relation between X and Y in Europe ($\rho_{x,y}$ = -0.41 in Europe).

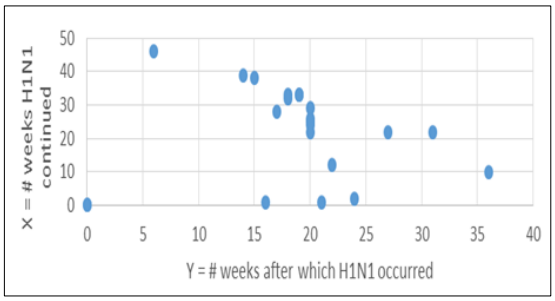

Figure 5: Regressive relation between X and Y in Americas ($\rho_{x,y}$ = -0.842 in Americas).

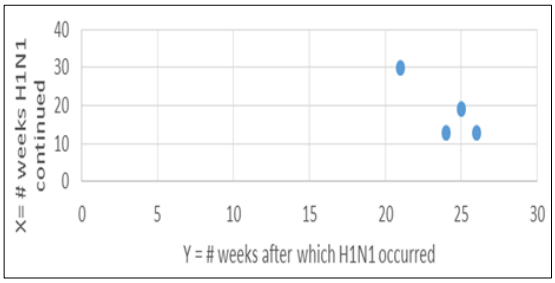

Figure 6: Regressive relation between X and Y in Oceanic ($\rho_{x,y}$ = -0.87 in Oceanic nations).

However, the conditional underlying model for the data on X, the number of weeks the H1N1 lasted, given that the H1N1 started in the y^{th} week is chosen to be a geometric distribution with $\lambda y > 0$ denoting the rate of H1N1's stopping. The joint PMF is then obtained.

From the joint PMF, the marginal PMF of X happened to be the regular geometric distribution with parameter $\lambda > 0$ without an involvement of y.

Hence, we obtained the conditional PMF of Y, the number of weeks to have elapsed for H1N1 to occur first time given that $X = x$ number of weeks H1N1 continued and it is named *Shanmugam's conditional geometric distribution* (SCGD) in this article as it has not been mentioned anywhere in the literature.

The marginal and conditional expected value and volatility (another name is variance) of the PMFs: $\Pr[Y = y|\theta]$, $\Pr[X = x|\theta,\phi,\lambda]$, $\Pr[X = x|Y = y,\lambda]$, and $\Pr[Y = y|X = x,\theta,\phi,\lambda]$ are obtained.

Using the marginal and conditional expected value and variances, the estimators $\hat{\theta}$. $\hat{\phi}$, $\hat{\lambda}$ of the model parameters and the estimator $\hat{Corr}[X,Y]$ of the correlation between X and Y, and the regressive projection $\hat{X}_{projection}$ of the number of weeks the H1N1 might continue on

occurring for a known $Y = y$, the number of elapsed weeks for H1N1 to have first occurred. These values are calculated, summarized, and displayed in the Table 2 and Figure 7, 8 & 9.

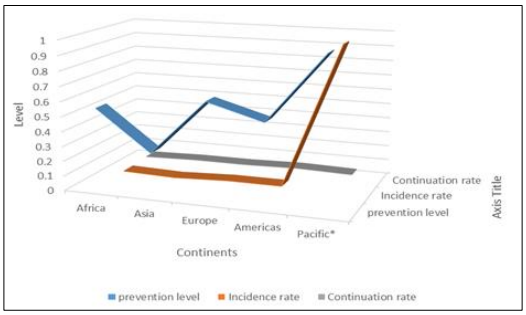

Figure 7: Incidence rate, prevention level, continuation rate of H1N1 (Comparison of Africa, Asia, Europe, Americas, Pacific).

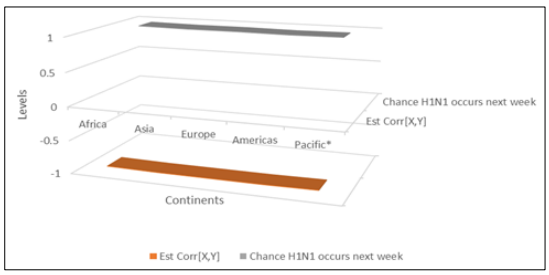

Figure 8: Estimated Probability of H1N1 in next week (Comparison of Africa, Asia, Europe, Americas, Pacific).

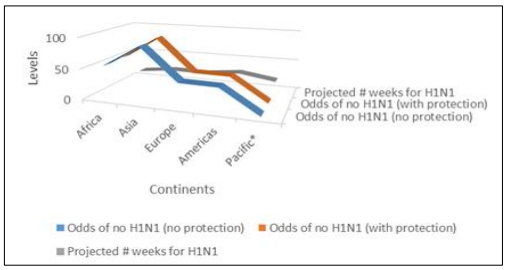

Figure 9: Projected # weeks H1N1 to occur, (Comparison of Africa, Asia, Europe, Americas, Pacific).

Surprisingly, Asia has the least deterrent against H1N1. The incidence rate and the continuation rate of H1N1 are closer to each other in other four continental regions except Oceanic (Figure 7). The survival probability for H1N1 to continue on in the next week once it has occurred in the 1st week of 2009 is consistently high in all five continental regions (Figure 8).

The *estimated negative correlation* between Y, the number of elapsed weeks for H1N1 to occur first and X, the number of weeks H1N1 to continue in without ending in 2009 is consistently high in all five continental regions with no exception (Figure 8).

The regressive projection $\hat{X}_{projection}$ of the number of weeks the H1N1 might continue on occurring for a known $Y = y$, the number of elapsed weeks for H1N1 to have first occurred ranges (Figure 9 or Table 2) from 16 weeks (in Africa) to 28 weeks (in Americas).

However, the odds for no H1N1 to occur (Figure 9 or Table 2) is lowest in Oceanic and highest in Asia in the absence as well as presence of deterrents to H1N1.

Table 1: Comparison of the results of 2009 H1N1 incidences (*IRRGD is not applicable because variance is smaller than then mean).

Continent → Result ↓	Africa	Asia	Europe	Americas	Pacific*
Preventive protection of H1N1: $\hat{\phi}$	0.55	0.28	0.647	0.544	0.99
H1N1 Incidence rate of H1N1 without protection: $\hat{\theta}$	0.02	0.01	0.026	0.026	0.99
H1N1 Continuation rate: $\hat{\lambda}$	0.03	0.04	0.04	0.049	0.04
Odds for no H1N1 incidence without protection: $\text{Odds}_{\theta, Y=0}$	54.4	89	37.81	36.9	0.01
Odds for no H1N1 incidence with protection: $\text{Odds}_{\theta,\phi, Y=0}$	55.2	91.5	38.37	37.76	0.06
Incidence rate of H1N1 with protection: $\hat{\theta}(1-\hat{\theta})^{(\frac{1}{\hat{\phi}}-1)}$	0.02	0.01	0.025	0.026	0.95
Given H1N1 occurred in 1st week, probability for it to occur in next week $S_{X\|y}(1)$	0.97	0.96	0.961	0.953	0.96
Estimated correlation: $\hat{Corr}[X, Y]$	-0.95	-0.96	-0.96	-0.96	-0.95
Projected number of weeks H1N1 would continue if it occurred in 1st week: $\hat{X}_{projection}$	16.3	22.8	21.84	27.92	18.8

Table 2: Summary for Africa, Asia, Europe, Americas, and Oceanic continents *(IRRGD is not suitable).

Description	Africa	Asia	Europe	Americas	Pacific*
Prevention level	0.55	0.28	0.647	0.544	0.99
Incidence rate	0.02	0.01	0.026	0.026	0.99
Continuation rate	0.03	0.04	0.04	0.049	0.04
Odds of no H1N1 (no protection)	54.4	89	37.81	36.9	0.01
Odds of no H1N1 (with protection)	55.2	91.5	38.37	37.76	0.06
Chance H1N1 occurs next week	0.97	0.96	0.961	0.953	0.96
Estimated Corr[X,Y]	-0.95	-0.96	-0.96	-0.96	-0.95
Projected # weeks for H1N1	16.3	22.8	21.84	27.92	18.8

COMMENTS AND CONCLUSIONS

From our analysis, we notice that no continental region is safe with respect to the occurrence of H1N1. The deterrent level ought to be strengthened. The healthcare administrators should perhaps look into ways and means of creating new deterrent measures. Such a conclusive recommendation to the worldwide healthcare administrators in all five continents has become possible with the help of the new and novel contents of this article primarily based on the seminal new distribution named as Shanmugam's Conditional Geometric Distribution (SCGD).

This new distribution is versatile enough for applications in other areas such as earthquake events, marketing, internet-security, and medical tourism. The statistics community ought to explore further research direction of SCGD.

ACKNOWLEDGEMENTS

The author thanks the School of Health Administration, College of Health Professions, and Texas State University, San Marcos, Texas for supporting with a sabbatical leave during spring semester, 2015 to complete the research work for this article.

Funding: The study was funded by a sabbatical leave from the Texas State University during spring 2015
Conflict of interest: None declared
Ethical approval: The author, Ramalingam Shanmugam takes responsibility for any error or opinion in the article

REFERENCES

1. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009;459(7250):1122-5.
2. Saito S, Minakami H, Nakai A, Unno N, Kubo T, Yoshimura Y. Outcomes of infants exposed to oseltamivir or zanamivir in utero during pandemic (H1N1) 2009. Am J Obstet Gynecol. 2013;209(2):130.e1-9.
3. Jain S, Kamimoto L, Bramley AM, Schmitz AM, Benoit SR, Louie J, et al. Hospitalized patients with 2009 H1N1 influenza in the United States, April-June 2009. N Engl J Med. 2009;361(20):1935-44.
4. Vijaykrishna D, Poon LL, Zhu HC, Ma SK, Li OT, Cheung CL, et al. Reassortment of pandemic H1N1/2009 influenza A virus in swine. Science. 2010;328(5985):1529.
5. Chiolero A, Paradis G, Rich B, Hanley JA. Assessing the relationship between the baseline value of a continuous variable and subsequent change over time. Frontiers Public Health. 2013;1:29.
6. Shanmugam R. Odds to quicken reporting already delayed cases: AIDS incidences are illustrated. Int J Nurs Res. 2013;4(1):1-13.
7. Maher M, Mountain L. The sensitivity of estimates of regression to the mean. Accid Analy Prev. 2009;41(4):861.

Cite this article as: Shanmugam R. Learning from an early start but late end epidemics via an incidence rate restricted bivariate distribution and data analysis. Int J Res Med Sci 2015;3:2181-9.