International Journal of Research in Medical Sciences
Shanmugam R. Int J Res Med Sci. 2014 May;2(2):462-467
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012

DOI: 10.5455/2320-6012.ijrms20140516

Research Article

Revelation of shrunken or stretched binomial dispersion and public perception of situations which might spread AIDS or HIV

Ramalingam Shanmugam*

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

Received: 03 January 2014
Accepted: 02 February 2014

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

ABSTRACT

Background: In 1985, the center for disease control coined the name: “Acquired Immune Deficiency Syndrome (AIDS)” to refer a deadly illness. The World Health Organization (WHO) estimated that about 33.4 million people were suffering with AIDS and two million people (including 330,000 children) died in 2009 alone in many parts of the world. A scary fact is that the public worry about situations which might spread AIDS according to reported survey result in Meulders et al. (2001). This article develops and illustrates an appropriate statistical methodology to understand the meanings of the data.
Methods: While the binomial model is a suitable underlying model for their responses, the data mean and dispersion violates the model’s required functional balance between them. This violation is called over-under dispersion. This article creates an innovative approach to assess whether the functional imbalance is too strong to reject the binomial model for the data. In a case of rejecting the model, what is a correct way of warning the public about the spreads of AIDS in a specified situation? This question is answered.
Results: In the survey data about how AIDS/HIV might spread according to fifty respondents in thirteen nations, the functional balance exists only in three cases: “needle”, “blood” and “sex” justifying using the usual binomial model (1). In all other seven cases: “glass”, “eating”, “object”, “toilet”, “hands”, “kissing”, and “care” of an AIDS or HIV patient, there is a significant imbalance between the dispersion and its functional equivalence in terms of the mean suggesting that the new binomial called imbalanced binomial distribution (6) of this article should be used. The statistical power of this methodology is indeed excellent and hence the practitioners should make use of it.
Conclusion: The new model called imbalanced binomial distribution (6) of this article is versatile enough to be useful in other research topics in the disciplines such as medicine, drug assessment, clinical trial outcomes, business, marketing, finance, economics, engineering and public health.

Keywords: Over/under dispersion, Ratio of statistics, p-value, Statistical power, Hypothesis test, Nuisance parameter

INTRODUCTION

On June 5, 1981, the Center for Disease Control (CDC) in USA publicized the biopsy of “5 young men with a rare pneumonia”. After failed immune system, their vital CD4+ cells were invaded by viruses, bacteria, fungi and parasites. The virus was detected by a Polymerase Chain Reaction (PCR). Later, CDC (1985) coined the word: “Acquired Immune Deficiency Syndrome (AIDS)” to refer this illness and announced that the world now faces “a new, deadly sexually transmitted disease” called AIDS. Who might have guessed in year 1981 that more than 45 million people would have died and another estimated 75 million people would have suffered worldwide with AIDS? The World Health Organization (WHO) estimated that about 33.4 million people were suffering with AIDS and two million people (including 330,000 children) died in 2009 alone in many parts of the world. The general public is frightened and has started worrying about every any or every situation in which the AIDS can be contracted. A survey data (Table 1) mirroring the public's perception on how AIDS might spread is made available in Meulders et al. (2001). A random sample of n = 50 persons from each of the thirteen countries was asked about the ten situations: eating a meal prepared by, handling an object from, drinking in a glass used by, sitting on a toilet used by, using the needle of, receiving blood from, shaking hands of, kissing mouth, having sex with, or taking care of an AIDS or HIV patient in which the AIDS might spread. Their responses were yes or no. The number, Y of "yes" for each situation is the entry in the Table 1. What do the data reveal? What is an appropriate model to adapt? These questions are the topics of discussion in this article.

Table 1: Y = # persons among in a sample of n = 50 in s = 13 nations agree that AIDS can be contracted in a specific activity source: Michel Meulders et al. (2001) journal of educational and behavioral statistics, 26 (2), 155-179.

Country (s=13)	Eating	Object	Glass	Toilet	Needle	Blood	Hand	Kissing	Sex	Care
France	12	10	24	25	49	50	7	24	50	24
Belgium	11	6	21	21	50	50	4	26	50	23
Netherlands	9	4	15	14	50	50	4	26	48	14
Germany	18	10	23	26	50	50	10	38	49	30
Italy	11	6	17	24	50	50	4	28	50	24
Luxemburg	5	5	21	20	50	50	6	25	50	15
Denmark	1	5	14	13	50	50	0	24	49	13
Ireland	15	11	22	27	50	50	8	36	50	16
Britain	11	8	26	14	49	49	3	34	49	19
Iceland	7	6	12	16	48	48	5	32	48	17
Greece	13	9	25	27	50	50	10	38	50	25
Spain	20	17	29	29	48	48	14	40	48	21
Portugal	29	24	32	38	48	49	24	45	49	34
Average, $\bar{y}$	12.4	9.3	21.6	23.6	49.4	49.5	7.6	32	49.2	21.7
Dispersion, s_y^2	54.8	31.6	34.8	54.37	0.76	0.60	37.4	49.2	0.69	40.1
Imbalance, $\hat{\phi}_{mle}$	0.71	0.61	0.48	0.63	0.11	0.14	0.71	0.62	0.04	0.53
p-value	0.00	0.00	0.20	0.00	0.98	0.97	0.00	0.03	0.99	0.08
Power	0.37	0.22	0.54	0.33	0.74	0.74	0.42	0.58	0.74	0.46
Prob ("yes")	0.25	0.19	0.43	0.47	0.99	0.99	0.15	0.64	0.99	0.43

NEED TO MODIFY BINOMIAL MODEL

Consider a scenario in which a random sample of *n* persons are selected and asked to express their opinion about a topic in terms of "yes" or "no". Let the probability for a response to be "yes" is $0 < p < 1$. This $\theta = \frac{p}{1-p} > 0$ means the *odds* of answering "yes" is . Assume that Y among the $n \geq 2$ answer "yes". Then, a natural choice to be underlying model for the random variable (RV) Y is a binomial model

$$\Pr(Y = y|\theta, n) = \frac{n!}{y!(n-y)!}\theta^{y}(1+\theta)^{-n}$$
$$n \geq 2; y = 0,1,2,...,n; 0 < \theta < \infty. \quad (1)$$

The mean and dispersion of the binomial model (1) are respectively in (2) and (3)

$$\mu = E(Y = y|\theta, n) = \frac{n\theta}{(1+\theta)} \quad (2)$$

and

$$\upsilon = Var(Y = y|\theta, n) = \mu(1 - \frac{\mu}{n}). \quad (3)$$

While their maximum likelihood estimates (MLE) are respectively $\hat{\mu}_{mle} = \bar{y}$ and $\hat{\upsilon}_{mle} = s_y^2$, the data should echo the balance requirement of the model (1) in the sense that $s_y^2 = \bar{y}(1 - \frac{\bar{y}}{n})$. The violation of this model requirement is recognized as over-under binomial dispersion in the data. A way to quantify the level of existing over-under dispersion in the data is by estimating the functional imbalanced measure

$$\hat{\phi}_{mle} = \frac{s_y^2 - \bar{y}(1-\frac{\bar{y}}{n})}{s_y^2 + \bar{y}(1-\frac{\bar{y}}{n})} . \qquad (4)$$

The parametric version of the functional imbalance (4) is rewritten, after algebraic simplifications, as in (5)

$$\theta = \frac{\mu - \upsilon + \phi(\mu+\upsilon)}{\upsilon(1-\phi)} . \qquad (5)$$

Now, fusing in the functional imbalance (5) into the binomial model, we notice that it becomes a bona fide model

$$\Pr(Y=y|\phi,\theta,n)$$
$$= \frac{n!}{y!(n-y)!}[\frac{(\mu-\upsilon)+\phi(\mu+\upsilon)}{\upsilon(1-\phi)}]^y$$
$$[\frac{\upsilon(1-\phi)}{\mu(1+\phi)}]^n$$
$$n \geq 2; y = 0,1,2,...,n;$$
$$-1<\phi<1; 0<\theta<\infty. \qquad (6)$$

The model (6) is not seen anywhere in the literature and hence, it is named here, for the first time, as functional imbalanced binomial (IBD). It is interesting to witness that when the functional imbalance is rectified to see balance (that is, $\phi = 0$), then the IBD (6) reduces to the balanced binomial model (1) as a special case. A further exploration of IBD reveals that its mean and dispersion are respectively in (7) and (8):

$$\mu_\phi = E(Y=y|\phi,\theta,n) = n[1-\frac{\upsilon(1-\phi)}{\mu(1+\phi)}] \qquad (7)$$

and

$$\upsilon_\phi = Var(Y=y|\phi,\theta,n)$$
$$= \frac{n\upsilon(1-\phi)}{\mu(1+\phi)}[1-\frac{\upsilon(1-\phi)}{\mu(1+\phi)}] \qquad (8)$$

When n = 1, the IBD (6) reduces to functional imbalanced Bernoulli's model in (9)

$$\Pr(Y=y|\phi,\theta,1)$$
$$= [1-\frac{\upsilon(1-\phi)}{\mu(1+\phi)}]^y[\frac{\upsilon(1-\phi)}{\mu(1+\phi)}]^{1-y}$$
$$y = 0,1; -1<\phi<1; 0<\theta<\infty. \qquad (9)$$

with the mean in (10)

$$E(Y=y|\phi,\theta,1) = [1-\frac{\upsilon(1-\phi)}{\mu(1+\phi)}] \qquad (10)$$

and in (11)

$$Var(Y=y|\phi,\theta,1) = \frac{\upsilon(1-\phi)}{\mu(1+\phi)}[1-\frac{\upsilon(1-\phi)}{\mu(1+\phi)}] \qquad (11)$$

To visualize them, let $E(Y=y|\phi,\theta,1)$, $\frac{\upsilon}{\mu}$ and ϕ are represented in z, x and y axis respectively. Then, the expectation (10) appears as in Figure 1,

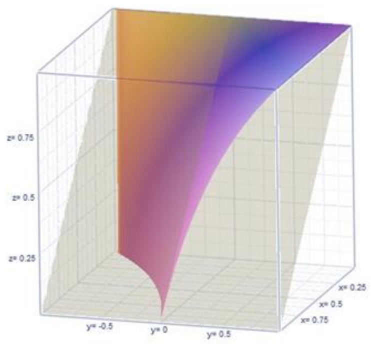

Figure 1: Expectation (10).

In which, the plane corresponds to the zero imbalance (that is, $\phi = 0$) and the bent configuration portrays the non-zero imbalance. The imbalance does alter the expectation.

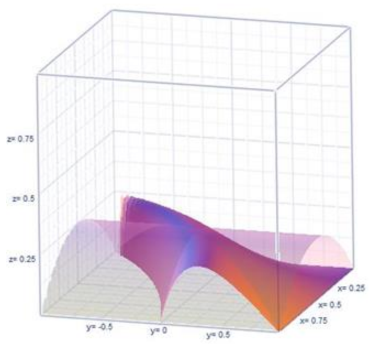

Figure 2: Dispersion (11).

The dispersion (11), on the contrary, appears like in Figure 2 in which the parabolic configuration corresponds to the zero imbalance (that is, $\phi = 0$) and the twisted configuration portrays the non-zero imbalance. The imbalance measure does indeed alter the dispersion.

In sensitive topics such as AIDS or HIV, it is not uncommon that all respondents exhibit the same answer "yes" or "no" to a question as an extremity depending on their knowledge level. In this framework, the probability for all respondents in the random sample to express "yes" is in (12)

$$\Pr(Y = n|\phi,\theta,n) = [1 - \frac{\upsilon(1-\phi)}{\mu(1+\phi)}]^n \quad (12)$$

which decreases to zero when the functional imbalance measure ϕ approaches $\frac{\upsilon-\mu}{\upsilon+\mu}$ and it happens in an extremely large sample size (that is, when $n \to \infty$). Likewise, the probability for all n in the random sample to express "no" is in (13)

$$\Pr(Y = 0|\phi,\theta,n) = [\frac{\upsilon(1-\phi)}{\mu(1+\phi)}]^n \quad (13)$$

which might decrease to zero when ϕ approaches 1.

The survival function $S(m|\phi,\theta,n) = \Pr(Y \ge m|\phi,\theta,n)$ of the IBD (6) can be expressed in term of cumulative F-model with 2m numerator and 2(n-m+1) denominator degrees of freedom as in (14)

$$S(m|\phi,\theta,n) = \Pr[F_{2m,2(n-m+1)} \le \{\frac{(n+1)}{m} - 1\}\{\frac{\mu-\upsilon+\phi(\mu+\upsilon)}{\upsilon(1-\phi)}\}] \quad (14)$$

after simplifications. In particular, when m = 1 in (11), the survival function, $S(1|\phi,\theta,n) = \Pr(Y \ge 1|\phi,\theta,n)$ portraying the probability that at least one respondent answers "yes" is in (15)

$$S(1|\phi,\theta,n) = \Pr[F_{2,2} \le n\{\frac{\mu-\upsilon+\phi(\mu+\upsilon)}{\upsilon(1-\phi)}\}]. \quad (15)$$

ESTIMATION OF PARAMETERS

Next, we need to estimate the parameters. For this purpose, consider a random sample $y_1, y_2, ..., y_s$ of size $s \ge 2$ from a IBD (6). Let $\bar{y}$ and s_y^2 be their sample mean and dispersion. The maximum likelihood estimators (MLE) of the parameters are in (16), (17) and (18):

$$\hat{\mu}_{mle} = \bar{y} \quad (16)$$

$$\hat{\upsilon}_{mle} = s_y^2 \quad (17)$$

and

$$\hat{\phi}_{mle} = \frac{s_y^2 - \bar{y}(1-\frac{\bar{y}}{n})}{s_y^2 + \bar{y}(1-\frac{\bar{y}}{n})}. \quad (18)$$

IS FUNCTIONAL IMBALANCE BETWEEN MEAN AND DISPERSION NEGLIGIBLE?

When in the data there is a significant tilt against the required functional balance between the mean and dispersion is significantly tilted, the usual binomial model (1) cannot be imposed on the data to interpret their findings. Instead, the imbalanced binomial model (6) is appropriate and should be considered. This process amounts to performing a statistical hypothesis testing. The powerful Neyman's $c(\alpha)$ technique is selected and exercised below. The $c(\alpha)$ technique is regression concept based. What is Neyman's *C* (α)? See Shanmugam (1992) for the procedural details. Following the principle, we obtain, after algebraic simplifications, the test statistic $T = \frac{s_y^2}{\bar{y}(1-\frac{\bar{y}}{n})}$ to test the null hypothesis $H_o : \phi = 0$ against the research hypothesis $H_1 : \phi \ne 0$. To perform the hypothesis test, we need to standardize the statistic. For this purpose, we next derive the expected value and dispersion of the test statistic T as follows. Using the formulas

$$E(\frac{U}{W}) \approx \frac{E(U)}{E(W)}\{1 + \frac{Var(W)}{[E(W)]^2} - \frac{Cov(U,W)}{E(U)E(W)}\},$$

$$E(h[W]) \approx h[E(W)],$$

$$Var(h[W]) \approx (\partial_w h[W])^2 Var(W)$$

and

$$Var(\frac{U}{W}) \approx [\frac{E(U)}{E(W)}]^2\{\frac{Var(U)}{[E(U)]^2} + \frac{Var(W)}{[E(W)]^2} - 2\frac{Cov(U,W)}{E(U)E(W)}\}$$

in Stuart and Ord (2009) with $U = s_y^2$ and $W = \bar{y}(1-\frac{\bar{y}}{n})$ where we find after algebraic simplifications that $E(T|\phi,\mu,\upsilon) = E[\frac{s_y^2}{\bar{y}(1-\frac{\bar{y}}{n})}|\phi,\mu,\upsilon] \approx 1$ and $Var(T|\phi,\upsilon) \approx \frac{(1+\upsilon_\phi)}{n\upsilon_\phi}$ where υ_ϕ is as defined in (8). Hence, the standardized statistic $Z = [\frac{s_y^2}{\bar{y}(1-\frac{\bar{y}}{n})} - 1]\sqrt{n(\frac{\upsilon_\phi}{1+\upsilon_\phi})}$ follows the standard normal model. It means that the p-value of rejecting the null hypothesis $H_o:\phi = 0$ in favor of the research hypothesis $H_1:\phi \neq 0$ is in (19)

$$p \approx 2\Pr[Z \geq \left|[\frac{s_y^2}{\bar{y}(1-\frac{\bar{y}}{n})} - 1]\sqrt{(\frac{s_y^2\{\bar{y}-s_y^2\}}{\bar{y}^2 + s_y^2\{\bar{y}-s_y^2\}})}\right|]. \quad (19)$$

When the null hypothesis is rejected at a α level and the true value of imbalance measure is known to be ϕ_1, the probability of accepting the true value ϕ_1 is called the statistical power and it is in (20)

$$power \approx \Pr[-\frac{z_{\alpha/2}}{\sqrt{n}}\left|\sqrt{\frac{[|\bar{y}^2 + s_y^2\{\bar{y}-s_y^2\}|][|s_y^2(1-\phi_1)\{\bar{y}(1+\phi_1) - s_y^2(1-\phi_1)\}|]}{\{|s_y^2\{\bar{y}-s_y^2\}|\}[\{\bar{y}(1+\phi_1)\}^2 + |s_y^2(1-\phi_1)\{\bar{y}(1+\phi_1)+s_y^2(1-\phi_1)\}|]}}\right| \leq Z \leq \frac{z_{\alpha/2}}{\sqrt{n}}\left|\sqrt{\frac{[|\bar{y}^2 + s_y^2\{\bar{y}-s_y^2\}|][|s_y^2(1-\phi_1)\{\bar{y}(1+\phi_1) - s_y^2(1-\phi_1)\}|]}{\{|s_y^2\{\bar{y}-s_y^2\}|\}[\{\bar{y}(1+\phi_1)\}^2 + |s_y^2(1-\phi_1)\{\bar{y}(1+\phi_1)+s_y^2(1-\phi_1)\}|]}}\right|] \quad (20)$$

RESULTS

In this section, we examine how the above methodology works in the survey data [Source: Michel Meulders et al. (2001)] about the AIDS/HIV contraction. A random sample of fifty (that is, n =50) persons in thirteen nations: Belgium, Britain, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxemburg, Netherlands, Portugal, and Spain. were selected and questioned on whether the ten activities: eating a meal prepared by, handling an object from, drinking in a glass used by, sitting on a toilet used by, using the needle of, receiving blood from, shaking hands of, kissing mouth, having sex with, or taking care of an AIDS or HIV patient might spread the AIDS/HIV illness. The number, Y among the fifty who answered "yes" are displayed in the Table 1.

The mean, $\bar{y}$ and dispersion, s_y^2 are calculated and shown in the Table 1. The probability for a respondent to answer "yes" for each one of the activities is estimated using (9) and displayed in the Table 1. By scanning their means and dispersions, it is easy to suspect that there might be an imbalance between the dispersion and its functional equivalent in terms of the mean. This imbalance fails the needed requirement for the usual binomial to be the underlying model. Using (4), the imbalance measure is calculated and displayed for each activity (see $\hat{\phi}_{mle}$ in the Table 1). Only in the case of "needle", "blood" and "sex", the imbalance is negligible and hence, the usual binomial is indeed the underlying model. This is confirmed by the higher p-value. In other words, the larger p-value indicates the likelihood for the null hypothesis $H_o:\phi = 0$ to be true. In the case of using the same "glass", the p-value is moderate for the null hypothesis $H_o:\phi = 0$ to be true. In other cases ("eating", "object", "toilet", "hands", "kissing", and "care" of an AIDS or HIV patient), the p-value is small enough to reject for the null hypothesis $H_o:\phi = 0$ meaning that $\hat{\phi}_{mle}$ is significant. In other words, for these six activities "eating", "object", "toilet", "hands", "kissing", and "care", the imbalanced binomial (6) is the appropriate underlying model.

Next, we examine the statistical power of accepting the true specific the research hypothesis $H_1:\phi_1 = 0.5$. It is calculated using (20) and displayed in the Table 1. The power is good across all the activities and above 50% in six out of ten activities.

CONCLUSION

Having witnessed that the new model called imbalanced binomial distribution (6) and the statistical methodology are too powerful and versatile not to be neglected in other areas of applications. What factors cause the imbalance to exist in the data? An answer to this question requires another methodology and it will be examined in future. More data with covariates are needed to create such a methodology.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. CDC. Revision of the case definition of acquired immunodeficiency syndrome for national reporting in the United States. Morb Mortal Weekly Rep. 1985;34:373-5.

2. Meulders M, Boeck PD, Mechelen IV, Gelman A, Maris E. Bayesian inference with probability matrix decomposition models. J Edu Behavioural Stat. 2001;26:153-79.
3. Shanmugam R. An inferential procedure for the Poisson intervention parameter. Biometrics. 1992;48:559-65.
4. Stuart A. K. Ord Kendal. In: Stuart A. K., J. Keith, eds. Kendall's Advanced Theory of Statistics. 1st ed. London U. K.: Griffin Publication; 2009: 356-359.

DOI: 10.5455/2320-6012.ijrms20140516
Cite this article as: Shanmugam R. Revelation of shrunken or stretched binomial dispersion and public perception of situations which might spread AIDS or HIV. Int J Res Med Sci 2014;2:462-7.