International Journal of Research in Medical Sciences
Shanmugam R Int J Res Med Sci. 2020 May;8(5):1606-1611
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012

DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20201904

Original Research Article

Distracted multinomial model for corona screening at entry ports

Ramalingam Shanmugam*

School of Health Administration, Texas State University, San Marcos, USA

Received: 15 March 2020
Accepted: 10 April 2020

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

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: On 24 January 2020, 1287 corona cases were noticed in Wuhan, China, causing 41 deaths. Its incubation period is at least 14 days. Now, this deadly virus has spread to other foreign countries. The prevalence of corona cases is changing daily. See www.who.org for daily reports. The corona cases are mystic and nightmare to the public, health professionals, and governing agencies globally.
Methods: The Center for Disease Control (CDC) compiled in their webpage (www.cdc.org) the number of confirmed, number of healthy, and the number of pending cases at the port of entries in United States of America (USA). These numbers are perhaps under-estimates because of inappropriate diagnostics and imprecise incubation period. To resolve the under estimating, this article, introduces a distracted multinomial model to refine the imprecise corona screening process and interpret the probability of detecting a corona case in US entry gates.
Results: An alternate expression (2) for the correlation between the corona ill and corona free cases at the USA ports of entry reveals that it was rising since 31st January 2020, reached its maximum on 5th February 2020, then declined to hit a bottom on 7th February 2020 only to rise again.
Conclusions: Most desirable is an accurate predictability of a traveller with the corona virus at the portal entry to minimize its spread. To make such prediction, a regression is necessary with involvement of covariates like age, body's immunity level, comorbidity, and precise understanding of its incubation period. The model in this article is the starting point for further future research work.

Keywords: Conditional prediction, Inappropriate diagnostic, Multinomial frequency trend, Model, Probability indices, Prevalence of corona cases, Under-estimate

INTRODUCTION

What is corona virus? It is witnessed to be an extremely acute respiratory deadly airborne disease whose cause or treatment remains unknown even at this time. The genesis of the name corona is traced to a Latin word corona, meaning crown or halo. The corona type illness can be traced to an earlier time in 1960 according to Habibzadeh and Stoneman.[1] Recently, during December 2019, the corona illness was first identified by health authorities in Wuhan, capital of Hubei province in China.[2,3] As of March 25, 2020, a total of 81,285 confirmed corona cases including 3,287 deaths in China. Among them, a total of 67,801 confirmed cases in Hubei Province (50006 cases in Wuhan) including 3169 death cases (2531 cases in Wuhan).[4] The corona virus (also recognized as COVID-19) originally spread to humans from animals but subsequently spread from a human to other humans.[5] The time between exposure and disease onset is typically 2 to 14 days.[6] The Symptoms may include fever, cough, and shortness of breath.[7] The Complications may include pneumonia and acute respiratory distress syndrome. There is no vaccine or specific antiviral treatment, with efforts typically aiming at managing symptoms and supportive therapy.[8] Hand washing is recommended to prevent the disease.[9] Anyone who is suspected of carrying the virus is advised to stay at home and monitor their health for two weeks, wear a mask, and seek medical advice by calling a doctor before visiting a clinic.[10]

Because the illness is seriously contagious, the travellers between the source locations in China and different parts of world are carriers of the corona virus, the corona incidences are noticed in several countries though in less numbers compared to the incidences within China. Since the first incidence of coronavirus (COVID-19) happened on Jan 20, 2020, in South Korea, the number of cases rapidly increased, resulting in 6284 cases including 42 deaths as of Mar 6, 2020. An empirical model on the reporting delay was speculated. The fatality rate was higher among males compared to females and it increased with older age. The implementation of social distancing was highly recommended.[11]

The World Health Organization (WHO) declared this outbreak of corona virus is a global emergency and recommended travel restrictions to and from Wuhan, Hubei, China. The train stations and airports within Wuhan, Hubei have installed body temperature checking booths to identify carriers and quarantine them. Many countries evacuated their citizens and diplomats from Hubei, China by chartered flights to their home nation. The USA is one of those countries. Around the world, especially in the United States of America (USA), traveller's from China are checked at the (sea, air) portal entries for the presence of corona virus. Neither an appropriate diagnostic test nor an effective treatment is available at the USA entry ports.[11,12]

On 30 January 2020, Center for Disease Control (CDC) witnessed the first USA citizen in Chicago, Illinois with corona virus as he returned home from Wuhan, China on 13 January 2020. The virus was confirmed in five other states in USA. As of 31 January 2020, the USA signed an exccutive order to deny entry to foreign nationals who travelled to China in the past two weeks. On the same day, the CDC declared a mandatory quarantine for 14 days for the 195 American evacuees from Wuhan, China. Even during the incubation periods (3 to 24 days, according to WHO), the virus is infectious from a human to another.[7]

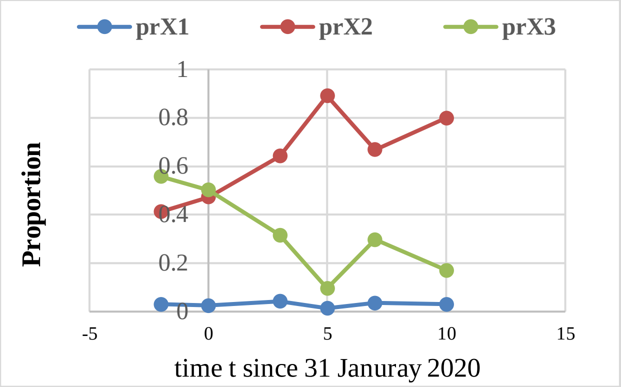
It is in this scenario of uncertainties, this article introduces an apt model, name it distracted multinomial and exercise it to understand the implications of the corona data (https://www.worldometers.info/usa-coronavirus for details).

As show in Table 1, a summary of the dates, the time, t since 31 January 2020, the numbers x_1, x_2, x_3 of positively, negatively, and pending cases respectively among those, $n = x_1 + x_2 + x_3$ tested cases in USA entry portals. The Figure 1 and Figure 2 depict respectively the chronologic (that is as of the time, t since the beginning 31 January 2020) proportions and index of dispersion (ID) in the three (that is, positive, negative, pending) categories. The ID value is the ratio of the variance to the mean. In the scenario with multinomial frequencies, the ID values are ought to be in the open interval [0,1] as it happened in this data.

Table 1: Corona at an entry in US (on 31 Jan, 2020, CDC announced that corona is a serious epidemic) (https://www.worldometers.info/usa-coronavirus).

Date	t	x_1	x_2	x_3	$n = x_1 + x_2 + x_3$
29/1/2020	-2	5	68	92	165
31/1/2020	0	6	114	121	241
3/2/2020	3	11	167	82	260
5/2/2020	5	11	706	76	793
7/2/2020	7	12	225	100	337
10/2/2020	10	12	318	68	398

x_1 = # Positive cases, x_2 = # negative cases, x_3 = # pending cases



X_1=Corona, X_1= No Corona, X_1=Pending

Figure 1: Chronic prevalence of corona cases.

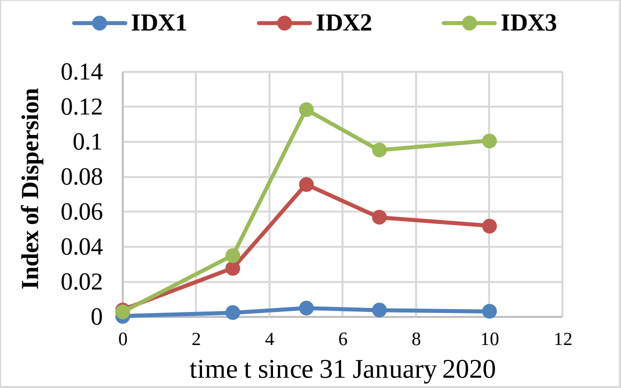


Figure 2: Index of dispersion of corona cases.

The outline of the articles is as follows. In Section2, author derive and assess properties of the distracted multinomial model in different sequential days since 31 January 2020, construct estimates of the model parameters, and hypothesis testing procedures. In Section 3, author illustrate the results for the USA data over the dates and make comparative judgements on the scrutiny of potential corona cases among the incoming travellers from Hubei, China. In Section 4, author make some conclusive remarks from data analytic points of view and their relevance for future dealings with respect to any other outbreak of infectious global diseases.

METHODS

In an incidence of infectious disease in which a proven diagnostic test is available, the probability for a randomly chosen suspect to get confirmed with the presence or absence of the disease is respectively **π**and 1-**π** because of the dichotomy. The corona virus turns the situation to a trichotomy because the practiced diagnostic(s) create a third category called pending category because of applying repeatedly several diagnostics. It is so because neither the cause nor the real incubation day(s) is mysterious. This reality defines the prevalences in the three categories as follows.

Let $\pi_1, \pi_2,$ and $(1-\pi_1-\pi_2)$ be the prevalences of corona ill, healthy from corona, and pending with respect to corona virus. Given that $n \geq 1$ travelers are subjected to a diagnostic test on a day at a port of entry to USA, then the probability for noticing the numbers x_1, x_2, and $x_3 = n - x_1 - x_2$ in the corona ill, healthy from corona, and pending categories is a multinomial type as in (1).

That is,

$$\Pr(x_1, x_2, n-x_1-x_2) =$$
$$\frac{n!}{x_1!x_2!(n-x_1-x_2)!}\pi_1^{x_1}\pi_2^{x_2}(1-\pi_1-\pi_2)^{n-x_1-x_2};$$
$$0<\pi_1,\pi_2<1; 0<\pi_1+\pi_2<1;$$
$$x_1, x_2 = 0,1,2,...,n; x_1+x_2 \leq n.$$

Because neither the causes of corona virus are not clear and the healthcare researchers have not formulated an effective/appropriate diagnostic test for corona infection, the health professionals at screening center of USA entry portals probably try one or another diagnostic in their list for detecting infectious corona virus. In the midst of uncertainty about the aptness of the diagnostic, let $0 \leq \phi < 1$ be the probability that a selected diagnostic test to try is *inappropriate*, then the prevalences get distorted to $\frac{\pi_1}{1-\phi}, \frac{\pi_2}{1-\phi}, 1-\frac{\pi_1+\pi_2}{1-\phi}$ with added restrictions,

$$0<\pi_1+\phi<1, 0<\pi_2+\phi<1, .$$
$$0<\pi_1+\pi_2+\phi<1$$

The underlying probability model for the outcomes now changes to,

$$\Pr(x_1, x_2, n-x_1-x_2) =$$
$$\frac{n!}{x_1!x_2!(n-x_1-x_2)!}(\frac{\pi_1}{1-\phi})^{x_1}(\frac{\pi_2}{1-\phi})^{x_2}$$
$$(1-\frac{\pi_1+\pi_2}{1-\phi})^{n-x_1-x_2}; \qquad (1)$$
$$0<\pi_1+\phi<1, 0<\pi_2+\phi<1,$$
$$0<\pi_1+\pi_2+\phi<1;$$
$$x_1, x_2 = 0,1,2,...,n; x_1+x_2 \leq n.$$

The model (1) is named distorted multinomial distribution. The third category is in fluid state unlike the other two categories because of the changing incupation time and/or opportunity to get scruitinized by another diagnostic. In other words, some cases may move on to eategory 1 (carona ill) or category 2 (healthy from corona) while a few might stay on in category 3 (pending still). The expected values, and variances of the distorted multinomial are interrelated. That is, $Var(X_i) = (1-\frac{\pi_i}{1-\phi})E(X_i)$ for $i = 1,2,3$ with an understanding that $X_3 = n - X_1 - X_2$. When the diagnostic for the classification is appropriate (that is with an appropriate disgnostic, $\phi = 0$), the variance-mean relationship becomes $Var(X_i) = (1-\pi_i)E(X_i)$, a well known identity in binomial/multinomial for $i = 1,2,3$. The relationship between the variance and expected value reveals that the index of dispersion (that is, the ratio of the variance over the expected value) is $ID_{X_i} = \{\frac{1}{1-\phi}\} - \{\frac{\phi+\pi_i}{1-\phi}\}$ for $i = 1,2,3$. Authors designate the quantity $CDI_\phi = \{\frac{1}{1-\phi}\} \in [1,\infty)$ as the aptness level (AL) of the diagnostic(s), where $\phi = \Pr[diagnostic\_is\_inappropriate]$. When the diagnostic is quite appropriate, note that the parameter, $\phi = 0$ and $CDI_{\phi=0} = 1$. Analoguously, authors designate the quantity $DC_{X_i} = \{\frac{\phi+\pi_i}{1-\phi}\}$ as the distortive classification *(DC) index* for category for $i = 1,2,3$.

When the diagnostic is quite appropriate (that is, $\phi = 0$), the classification index is for $i = 1,2,3$ is $DC_{X_i} = \pi_i$. Otherwise, when an inappropriate diagnostic is exercised, an amount of distortiveness occurs and the deviation amount from an ideal classification index π_i is $-(\frac{\phi}{1-\phi})\pi_i - \frac{\phi}{1-\phi}$. An interpretation of $-\frac{\phi}{1-\phi}$ in the deviation in the distortiveness is the difference $CDI_{\phi=0} - CDI_\phi$. Notice that this difference is not only the initial amount but also the rate of change in the deviation.

Also, let us focus on the pending category. Note $Var(X_3) = Var(n - X_1 - X_2) = Var(X_1 + X_2)$ which is equal to,

$$\frac{Var(X_1) + Var(X_2)}{+2\rho_{X_1,X_2}\sqrt{Var(X_1)Var(X_2)}},$$

where ρ_{X_1,X_2} is the correlation between the number of *corona ill* and healthy (corona ill free) cases. An easily recognizable is the expected value of the pending category is

$$E(X_3) = n - E(X_1) - E(X_2) = n(1 - \frac{\pi_1 + \pi_2}{1-\phi}).$$

Hence, the index of dipsersion for the pending category is a weighted average of the indices of dispersion for the corona ill category and the corona virus free healthy category with their expected values as weights. That is

$$ID_{X_3} = \frac{\begin{array}{c} ID_{X_1}E(X_1) + ID_{X_2}E(X_2) \\ +2\rho_{X_1,X_2}\sqrt{ID_{X_1}E(X_1)ID_{X_2}E(X_2)} \end{array}}{\{n - E(X_1) - E(X_2)\}}. \quad (2)$$

An implication of the above relationship is summarized in theorem 1 below for the correlation between the corona ill and corona virus free healthy categories.

Theorem 1: The correlation, ρ_{X_1,X_2} between the corona ill cases, X_1 and the corona virus free healthy cases, X_2 is

$$\rho_{X_1,X_2} = [\frac{n - E(X_1) - E(X_2)}{n}]ID_{X_3},$$

where $ID_{X_3} = 1 - \frac{\pi_3}{1-\phi} = (\frac{\pi_1 + \pi_2}{1-\phi})$, $Var(X_i)$ and $E(X_i)$ are indices of dispersion of pending category, variance and expected values for $i = 1, 2$.

An implication of the theorem is that the correlation ρ_{X_1,X_2} increases proportionally as the index of dispersion ID_{X_3} increases with the proportionality equal to $\frac{n - E(X_1) - E(X_2)}{n}$. Next, author have to estimate the model parameters. It is quite clear that the maximum likelihood estimate (MLE) is most efficient.[15] The MLE has a virtue that the MLE of a function of the parameters is the function of the MLE of the parameters.[14] To be specific, the MLEs of $\hat{ID}_{X_3} = (1 - \frac{\hat{\pi}_3}{1-\hat{\phi}}) = \frac{n - \bar{x}_3}{n}$, $V\hat{a}r(X_i) = s^2_{x_i}$, and $\hat{E}(X_i) = \bar{x}_i$, where $s^2_{x_i}$ and $\bar{x}_i$ denote respectively sample variance and mean for category $i = 1, 2$. The MLE of the inappropriate level of the diagnostic is $\hat{\phi} = 1 - \{\frac{\hat{\pi}_1}{\hat{\pi}_2}\}$. Sunstituting these MLEs in the alternate expression for the correlation ρ_{X_1,X_2} in Theorem 1 and that is $\hat{\rho}_{X_1,X_2} = [\frac{n - \bar{x}_1 - \bar{x}_2}{n}]ID_{X_3}$. These MLEs for the USA corona data in Table 1 are displayed in Table 2 as of the closing dates.

Table 2: Correlation, index of dispersion on the closing date for the USA data on corona cases.

Closing day	Total cases	$\bar{x}_1$	$\bar{x}_2$	$s^2_{x_1}$	$s^2_{x_2}$	$(\frac{\hat{\pi}_1}{1-\hat{\phi}})$	$(\frac{\hat{\pi}_2}{1-\hat{\phi}})$	$\hat{\phi}$	$\hat{ID}_{x_3}$	$\hat{\rho}_{x_1,x_2}$
31/1/2020	241	5.5	91	0.5	1058	0.02	0.38	0.94	0.56	0.33
3/2/2020	260	7.3	116.3	10.3	2454.3	0.03	0.45	0.93	0.62	0.33
5/2/2020	793	8.2	263.7	10.2	88562.9	0.01	0.33	0.97	0.88	0.58
7/2/2020	337	9	256	10.5	66722.5	0.04	0.76	0.96	0.72	0.15
10/2/2020	398	9.5	266.3	9.9	54018.6	0.02	0.67	0.96	0.77	0.24

RESULTS

In this section, author illustrate the alternate expression (2) for the correlation between the corona ill and corona ill free cases under the distorted multinomial model for the USA data in Table 1. The average number of corona cases, $\bar{x}_1$ is the observed number of corona cases over the total cases.

The average number of healthy cases, $\bar{x}_2$ is the obsedrved number of healthy cases over the total cases, n. The sample variances $s^2_{x_1}$ and $s^2_{x_2}$ are based on the observed corona and healthy cases as of the chosen closing day. The mle, $(\frac{\hat{\pi}_i}{1-\hat{\phi}})$ are the ratio $\frac{\bar{x}_i}{n}$ for $i = 1, 2$. The mle, $\hat{\phi} = 1 - \frac{\hat{\pi}_1}{\hat{\pi}_2}$ portrays the estimated level of inappropriate level of the employed diagnostic to identify the corona ill cases. The mle of the index of dispersion is done using $\hat{ID}_{X_3} = (1 - \frac{\hat{\pi}_3}{1-\hat{\phi}}) = \frac{n - \bar{x}_3}{n}$. The mle of the correlation is done employing the core result, $\hat{\rho}_{X_1,X_2} = [\frac{n-\bar{x}_1-\bar{x}_2}{n}]\hat{ID}_{X_3}$ of the theorem 1 in this article. The proportionality between the estimated correlation, $\hat{\rho}_{x_1,x_2}$ and the index of dispersion, $\hat{ID}_{x_3}$ of the pending cases is displayed in figure.

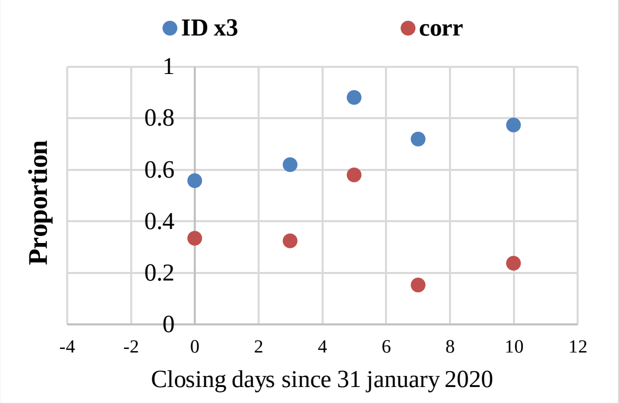


Figure 3: Proportionality of $\hat{\rho}_{x_1,x_2}$ versus and $\hat{ID}_{x_3}$.

The ASC was more in males bilaterally than in females. The difference was statistically significant on right side (p=0.00). The ASS was more in females than in males on right and vice versa on the left side.

DISCUSSION

The predictability of a traveler into a category of corona ill is desirable in an entry portal. To make such prediction, a regression would help.[13] To construct such a regression, data on covariates are necessary and the data are not in a public domain for several reasons of privacy and practicality. However, the developments in this article about the underlying probability model for the chance mechanism for the outcomes with reference to travelors with corona virus remain the foundation for further research work.

A multivariate logistic curve for the number of deaths in several countries due to corona virus is feasible based on the distorted multinomial model of this article. These require a comprehensive data collection apparatus and it can be implemented perhaps by the WHO in future pandemics. The knowledge saves life.

It will be of great practical importance to understand the cross-country spread of the corona pandemic for the sake of complete global prevention. Also, the cross-regional connections among cities as well as between urban and rural sections within a country are worthwhile in the pursuit of stopping the deadly corona virus.

A space-time correlation analysis of the distance from the epicenter (that is Wuhan in China) and the other global locations the number of corona cases will reveal a lot more the role played by the travellers and the gravity law in the social dynamics sense.

ACKNOWLEDGEMENTS

The author thanks the School of Health Administration, College of Health Professions, Texas State University to complete the research work for this article.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Habibzadeh P, Stoneman EK. The novel coronavirus: a bird's eye view. Int J Occup Environ Med. 2020 Mar 27;11(2 April):1921-65.
2. Zhang Y. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) – China, 2020. Available at: http://weekly.chinacdc.cn, China CDC Weekly. 2.
3. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar;579(7798):270-3.
4. Heymann DL, Shindo N. COVID-19: what is next for public health?. Lancet. 2020 Feb 22;395(10224):542-5.
5. 2019 Novel Coronavirus. Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/coronavirus/2019-ncov/about/symptoms.html. Accessed 11 February 2020.
6. Symptoms of Novel Coronavirus (2019-nCoV) | CDC, U.S. Centers for Disease Control and Prevention (CDC). Available at: https://www.cdc.gov/coronavirus/2019-ncov/about/symptoms.html. Accessed 10 February 2020.
7. Prevention and Treatment, Centers for Disease Control and Prevention (CDC). Available at: https://www.cdc.gov/coronavirus/about/prevention.html. https://web.archive.org/web/2019115193934/https://www.cdc.gov/coronavirus/about/prevention.html. Accessed 9 August 2019.
8. Prevention and Treatment. Centers for Disease Control and Prevention (CDC). Available at: https://www.cdc.gov/coronavirus/about/prevention.html. Accessed 9 August 2019.
9. Updates on Wuhan Coronavirus (2019-nCoV) Local Situation, Available at: https://www.moh.gov.sg/2019-ncov-wuhan.moh.gov.sg.
10. National Health Commission (NCP). National Health Commission of the People's Republic of

China. Available at: http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm#NHCFeb10. Accessed 10 February 2020.

11. Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. Int J Infect Dis. 2020 Mar 18;93:339-44.
12. Wilder-Smith A, Freedman DO. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J Travel Medi. 2020 Mar;27(2):20.
13. Dell'Omodarme M, Prati MC. The probability of failing in detecting an infectious disease at entry points into a country. Statist Medi. 2005 Sep 15;24(17):2669-79.
14. Shanmugam R, Chattamvelli R. Statistics for scientists and engineers. John Wiley Press, Hoboken, NJ; 2016.
15. Stuart A, Ord JK. Kendall's Advanced Theory of Statistics. In: Kendall MG, Stuart A, Ord JK. Volume 1. Oxford University Press, London. 2015.

Cite this article as: Shanmugam R. Distracted multinomial model for corona screening at entry ports. Int J Res Med Sci 2020;8:1606-11.