HETEROGENEITY VERSUS THE COVID-19 PANDEMIC

Ramalingam Shanmugam[1], Gerald Ledlow[2] and Karan P. Singh[3]

[1]Professor of Statistics and Honorary Professor of International Studies

School of Health Administration, Texas State University

San Marcos, TX 78666, USA

Email: rs25@txstate.edu

ORCID number 0000-0002-3388-1014

[2] Professor and Dean, Department of Healthcare Policy, Economics and Management

School of Community and Rural Health

The University of Texas Health Sciences Center at Tyler

Tyler, TX 11937, USA

Email: gerald.ledlow@uthct.edu

[3] Professor and Chair, Department of Epidemiology and Biostatistics

The University of Texas Health Sciences Center at Tyler

Tyler, TX 11937, USA

Email: karan.singh@uthct.edu

ORCID number 0000-0002-1841-846X

Abstract

In this paper, heterogeneity is formally defined, and its properties are explored. We define and distinguish observable versus non-observable heterogeneity. It is proposed that *heterogeneity* among the vulnerable is a significant factor in the contagion impact of COVID-19, as demonstrated with incidence rates on a Diamond Princess Cruise ship in February 2020. Given the nature of the disease, its heterogeneity and human social norms, pre-voyage and post-voyage quick testing procedures may become the new standard for cruise ship passengers and crew. The technological advances in testing available today would facilitate more humanistic treatment as compared to more archaic quarantine and isolation practices for all onboard ship. With quick testing, identification of those infected and thus not allowed to embark on a cruise or quarantining those disembarking and other mitigation strategies, the popular cruise adventure could be available safely again. Whatever the procedures implemented, the methodological purpose of this study should add valuable insight in the modeling of disease and specifically, the COVID-19 virus.

***Key Words*:** Observed homogeneity; non-observed homogeneity; over dispersion; under dispersion; Poisson distribution; binomial distribution; Tango's test statistic.

1. INTRODUCTION

In the literature, the term *heterogeneity* echoes differently in various contexts. What is heterogeneity or its antonym, homogeneity? Its root word lies in Greek "*heterogenes*" meaning different. In epidemiology or statistics disciplines, the word heterogeneity is popularly commented to exist when the variance is large. In insurance applications, for an example, the premium is assessed more if the insurer is in a heterogeneous group with high hazard proneness (Spreeuw, 1999). Should a large (small)

variance be indicative of heterogeneity (homogeneity)? Interesting discussions are given for heterogeneity in Ecochard (2006); in healthcare disciplines, heterogeneity is referred to as different outcomes among patients. Should the heterogeneity be connected to only a non-observable hidden trait as done in genetics? Does heterogeneity refer dissimilar attributes across the subgroups of the population itself even before sampling? Is heterogeneity really pointing to the non-identical nature in a random sample or population? Should heterogeneity imply a shifting entity? In genetic studies, several authors refer to genetic heterogeneity as rather too difficult to ascertain. What do they really mean? If alleles in more than one locus exhibit susceptibility to a disease, there is a need to track the loci to infer their heterogeneity. So, in a sense, the application of heterogeneity is really a discussion of an opposite of similarity across loci. The reader is referred to Elston et al. (2003, pages 3404-344) for details. Hope and Norris (2013) attempted to determine how heterogeneity played a role in judgements in the context of crime victimization. Hence, what really is heterogeneity? A formal definition of heterogeneity is constructed later in the article, then, its properties are explored and itemized.

However, in the epidemiology literature, using a random sample $y_1, y_2, ..., y_n$ from a population whose main parameter is θ , when the null hypothesis $H_o : \theta_1 = \theta_2 = ... = \theta_n$ is tested, it is named the *homogeneity test.* This suggests that heterogeneity is really all about a shifting population. This creates more confusion. Is the source of such confusion with respect to heterogeneity its ill communication? It is evident that there is a lack of a clear definition of heterogeneity given by Hunink et al. (2018, Chapter 12) for details. Neither the *Encyclopedia of Statistical Sciences n*or the *Encyclopedia of Biostatistics* has even an entry, as if it is not pertinent in statistical disciplines.

One comes across different types of data in epidemiologic studies. Drawing data from a binomial population is one of them, and the data should possess an under dispersion (i.e., variance of the binomial distribution is smaller than its mean). From a Poisson population, the drawn random sample ought to

reflect equality between the mean and variance. When the main (incidence rate) parameter of a Poisson chance mechanism is stochastically transient, the unconditional observation of the random variable convolutes to an inverse binomial model (Ross, 2002). The inverse binomial distribution is known to attest that the variance is larger than its mean (Stuart and Ord, 2015, for details). Consequently, a comparison between the mean and variance characterizes only which type binomial, Poisson, or inverse binomial possesses the underlying chance mechanism we are sampling from but does not inform anything about heterogeneity.

With details about the probabilistic patterns among coronavirus confirmed, recovered, or cured individuals and those that succumb as fatalities/deaths in the thirty-two states/territories of India are given by Shanmugam (2020). To track the confusion with respect to heterogeneity, let us consider the data given in Table 1 (Mizumoto and Chowell, 2020), describing the spread of COVID-19 among the voyagers in a Diamond Princess Cruise ship, during the month of February 2020. The random variables Y_1, Y_2, and Y_3 denote, respectively, the number of COVID-19 cases, the number of asymptomatic cases and the number of symptomatic cases among them in time (date). Under a given COVID-19's prevalence rate, $\lambda > 0$, the number Y_1 perhaps follows a Poisson probability pattern. For a given number of COVID-19 cases in a date, the number Y_2 perhaps follows a binomial probability pattern with parameters (y_1, p), where $0<p<1$ denotes the chance for a COVID-19 case to exhibit no symptoms. Naturally, the number Y_3 should follow a binomial probability pattern with parameters $(y_1, 1-p)$. There is an implicitness between Y_2 and Y_3, in the sense that $Y_2 + Y_3 = Y_1$. There are three-time oriented groups of COVID-19 incidences in Table 1. Is there an observable heterogeneity among the three groups? If so, is it due to a non-observable (parametric) heterogeneity? How do we define and distinguish observable versus non-observable heterogeneity? A literature search in epidemiology and/or biostatistics does not provide an answer to this question.

It is evident that the average of COVID-19 cases is an estimate of COVID-19's prevalence rate (i.e., $\hat{\lambda}$ in Table 1). Their estimates impress that the prevalence rate is transient, not constant across every pair of two-day duration dyads. The Poisson population from which the COVID-19 cases are drawn ought to have been dynamic, implying the existence of a Poisson heterogeneity. How do we define and/or capture the heterogeneity level? This is the theme and purpose in this research article.

Likewise, given that a fixed number, y_1 of COVID-19 cases has occurred, a part of them might be asymptomatic cases, y_2 and the remaining are symptomatic cases, y_3. That is, y_2 and y_3 are complementary but $y_2 + y_3 = y_1$. Is there heterogeneity in each of the two sub-binomial populations, whether there is a heterogeneity in y_1? How should each *binomial heterogeneity* be defined and computed? In other words, is binomial heterogeneity different from that of Poisson heterogeneity? If so, what are the differences? A literature search in epidemiology and/or biostatistics offers no help to prove either the existence or absence of binomial heterogeneity in the data for y_2 or y_3 in Table 1. Hence, we continue probing matters with respect to heterogeneity.

The concept of heterogeneity seems to have escaped the researchers and epidemiologists' scrutiny for a long time. It is time well spent and worthwhile to revive an interest in the construct of heterogeneity, and that is exactly what this article is trying to accomplish. Hence, we first define and construct an approach for the idea of heterogeneity. To be specific, we first discuss Poisson heterogeneity and then take up binomial heterogeneity. Maybe our research direction about heterogeneity is, perhaps, pioneering. However, we believe that our approach is easily extendable for many other similar methodological setups. We illustrate our definition and all derived expressions for heterogeneity using COVID-19's data pertaining to the Diamond Princess Cruise ship, Yokohama, 2020 as displayed in Table 1.

2. POISSON AND BIONOMIAL HETEROGENEITIES

Applied epidemiologists emphasize that heterogeneity is of paramount importance in extracting and interpreting data evidence. Many data analysts are convinced that an unrecognized heterogeneity leads to a biased inference. To begin with, what is heterogeneity? It is a factor causing non-similarities. If so, how many sources are there? We contemplate that there are two sources for heterogeneity to exist. One source ought to be from the drawn random sample of observations: $y_1, y_2, ..., y_n$, which we recognize as *observable heterogeneity*. Would the sampling variability, $Var[f(y_1, y_2, ..., y_n|\theta)]$ for a selected statistic $f(y_1, y_2, ..., y_n|\theta)$ express the observable heterogeneity? Another source is manifested in non-observable parameter, θ of the chance mechanism, which we recognize as *non-observable heterogeneity*. Would a non-uniform stochastic pattern of θ be indicative of the non-observable homogeneity? If the chance mechanism perversely selects a probability density function (pdf) for θ, how would it manifest itself to portray the non-observable heterogeneity? Both observable and non-observable heterogeneity together ought to be involved to make any definition of heterogeneity complete. If so, how do we integrate them? Often, under/over-dispersion is confused with heterogeneity. It seems that the over/under dispersion is precipitated by heterogeneity but not the other way. It is not obvious or proven so far in the epidemiology literature on whether the converse is true. We focus only on Poisson and binomial populations to address heterogeneity, and these arguments can be repeated for other populations considering similar methods.

2.1. POISSON HETEROGENEITY

Recall that the random integer, Y_1 denoting the number of COVID-19 cases in a place (like the Diamond Princess cruise ship) at a time (like February, 2020) is a Poisson random variable with a specified prevalence rate, $\lambda > 0$. That is, the conditional probability of observing y_1 number of COVID-

19 cases under a prevalence rate $\lambda > 0$ is $\Pr[Y_1 = y_1 | \lambda] = e^{-\lambda}\lambda^{y_1} / y_1!; y_1 = 0,1,2,...; \lambda > 0$ with its expected number $E[Y_1 | \lambda] = \lambda$ and variability $Var[Y_1 | \lambda] = E[Y_1 | \lambda]$. The reader is referred to Rajan and Shanmugam (2020) for detailed derivations of the Poisson mean and variance. The prevalence parameter λ itself is crucial in our discussions. The Poisson variability cannot be heterogeneity, because the expected value also changes when the variability changes due to their inter-relatedness. Realize that no two individuals on the ship are assumed to have the same level of susceptibility to the COVID-19 virus. It is reasonable to imagine that the prevalence levels follow a conjugate, stochastic gamma distribution. The so-called conjugate prior knowledge in the Bayesian framework smooths the statistical analytic process. It is known that the conjugate prior for the Poisson distribution is gamma, whose pdf is

$$c(\lambda | \alpha, \beta) d\lambda = e^{-(\alpha\lambda)}(\alpha\lambda)^{\beta-1} d(\alpha\lambda) / \Gamma(\beta); \alpha > 0; \beta > 0, \quad (1)$$

with an average. $E(\lambda | \alpha, \beta) = \frac{\beta}{\alpha}$ and variability $Var(\lambda | \alpha, \beta) = E(\lambda | \alpha, \beta) / \alpha$, where the parameters α and β are recognized as *hyper-parameters* (Rajan and Shanmugam, 2020). Notice that the hyper parameter $\alpha > 0$ causes the variability in the COVID-19's prevalence rate to fluctuate up or down, and, hence, you would anticipate the heterogeneity to involve the hyperparameter α. But the question is how?

We assume that the probability of observing a non-negative COVID-19 case, y_1 is a Poisson under a stable sampling population $\Pr(Y_1 | \lambda)$ with an expected number $E(Y_1 | \lambda) = \lambda$ and a variability $Var(Y_1 | \lambda) = E(Y_1 | \lambda)$. With replications, the observable heterogeneity should become estimable. That is to mention, the maximum likelihood estimate (MLE) of the COVID-19 prevalence rate is the average number, $\bar{y}_1$, of the observations. To discuss the non-observable heterogeneity, we need to integrate its

conjugate prior $c(\lambda|\alpha,\beta)$ for the non-observable λ with the likelihood $\Pr(Y_1|\lambda)$ and it results in an update and it is called posterior pdf for λ . The expressions for non-observable heterogeneity, observable heterogeneity and other expressions are given in Appendix I.

2.2. BINOMIAL HETEROGENEITY

In this section, we explore heterogeneity for two sub-binomial processes emanating from a Poisson process. The asymptomatic number, Y_2 and symptomatic number, Y_3 of COVID-19 cases are two branching binomial random numbers out of the Poisson random number, $Y_1 = 0,1,2,...;$ of COVID-19 cases. These two split random variables are complementary of each other in the sense that $Y_2 + Y_3 = Y_1$. Then, what are the underlying model for Y_2 and for Y_3? Are they correlated random variables? If so, what is their correlation? These are pursued in this section.

Let *I* be an indicator random variable defined as: $I_i = 1$ for a COVID-19 case to be asymptomatic with a probability, $0<p<1$ and $I_i = 0$ for the case to be symptomatic with a probability, $0<1-p<1$. Then, for a fixed y_1, the random variable, $Y_2 = \sum_{i=1}^{Y_1} I_i$ follows a binomial probability distribution with parameters (y_1, p). Likewise, for a fixed y_1, the random variable, $Y_3 = y_1 - Y_2$ follows a complementary binomial distribution with parameters $(y_1, 1-p)$.That is,

$$\Pr(Y_2 = y_2|y_1, p) = \binom{y_1}{y_2} p^{y_2}(1-p)^{y_1-y_2}; y_2 = 0,1,2,.....,y_1; 0<p<1 \quad (2)$$

and

$$\Pr(Y_3 = y_3|y_1, p) = \binom{y_1}{y_3}(1-p)^{y_3}\, p^{y_1-y_3}; y_3 = 0,1,2,.....,y_1; 0<1-p<1 \quad (3)$$

The expressions for non-observable heterogeneity, observable heterogeneity and other expressions are given in Appendix II.

3. TANGO INDEX

Lastly, we develop the *Tango index* and its significance level over the time period. Tango (1984) proposed an index to detect disease clusters in grouped data. This index received considerable attention in the literature. Following the line of thinking in Tango (1984), we could next assess the MLEs of several entities we estimated and displayed in Tables 1, 2, and 3. There are three groups of duration. Group 1 consists of the 15th and 16th of February 2020. Group 2 includes data for 17th and 18th of February 2020. Group 3 contains data of 19th and 20th of February 2020. Two independent contrasts among the three groups are feasible. In an arbitrary style, we select to compare Group 1 with Group 2 and then Group 2 with Group 3. For this purpose, we formulate a contrast matrix

$$A_{3x3} = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad (4)$$

where the third column of the matrix needs no explanation. The Tango's statistic $T = \underline{r}' A \underline{r}$ follows a chi-square distribution with $v = 2$ degrees of freedom (df), where $\underline{r}_{x3}^{'}$ is a row vector of the MLE of a chosen entity in our analytic results in Table 1 or Table 2 or Table 3. For an example, let $\underline{r}' = (68.5, 93.5, 46)$ for the MLE of the COVID-19 prevalence rate, λ in the groups. Then, the Tango's test statistic is $T = 422.25$ with $v = 2$ df and $p - value = 2.03975E - 92$. Likewise, the Tango's test statistic value and its p-value are calculated and displayed in Table 4 for other entities.

4. ILLUSTRATING USING COVID-19 DATA OF THE DIAMOND PRINCESS CRUISE SHIP

197 In this section we illustrate all the concepts and expressions of Section 2. Let us consider the
198 COVID-19 data in Table 1 for the Diamond Princess Cruise Ship, 2020. The Diamond Princess is
199 a cruise ship registered in Britain and operated across the globe. During a cruise that began on 20
200 January 2020, positive cases of COVID-19 linked to the pandemic were confirmed on the ship in
201 February 2020. Over 700 people out of 3,711 became infected (567 out of 2,666 passengers and 145 out
202 of 1,045 crew), and 14 passengers died. To be specific, on the 15th of February 2020, 67 people were
203 infected, on the 16th of February 2020, 70 people were infected, on the 17th of February 2020, there were
204 99 COVID-19 cases, on the 18th of February, another 88 cases were confirmed. The U.S. government
205 initially asked Japan to keep the passengers and crew members on board the ship for 14 days. The U.S.
206 government, however, later decided to bring them to an Air Force base in California and a base in San
207 Antonio, Texas.

208 For each specified day in the first column in Table 1, the estimate of COVID-19's prevalence rate
209 and its variance are calculated using expressions $\hat{\lambda} = \overline{y}_1$ and $\hat{Var}(Y_1 | \lambda) = s^2_{y_1}$. Both the prevalence and its
210 variability increased and then decreased over the days. However, their correlation, $\hat{\rho}_{Y_2, Y_3}$ is calculated
211 using the observed numbers on y_2 and y_3 for each day (see in Table 2) and the estimated correlations had
212 been stable over the days. Substituting $\hat{\lambda} = \overline{y}_1$ and $\hat{Var}(Y_1 | \lambda) = s^2_{y_1}$ in the expression

213 $$\hat{H}_\lambda = \frac{\hat{\lambda}}{\hat{\lambda} + \hat{Var}(Y_1 | \lambda)}, \qquad (5)$$

214 we obtained the non-observable heterogeneity and displayed in Table 2. The non-observable Poisson
215 heterogeneity for y_1 was high on the beginning day, came down later, and then increased. Using $\hat{\lambda} = \overline{y}_1$ and
216 $\hat{Var}(Y_1 | \lambda) = s^2_{y_1}$ in the expression

$$\hat{H}_{\bar{y}_1} = [1 + \frac{(1+\frac{\bar{\lambda}}{s_\lambda^2})}{6}]^{-1}, \quad (6),$$

we obtained the observable heterogeneity and displayed in Table 2. The observable Poisson heterogeneity was low on the first day, increased and then decreased. Note in Table 2 that the observable and non-observable Poisson heterogeneities are inversely proportional. In other words, the estimate of the shape and scale parameter in the Bayesian approach are respectively $\hat{\alpha} = \frac{\bar{\lambda}^2}{s_\lambda^2}$ and $\hat{\beta} = \frac{\bar{\lambda}}{s_\lambda^2}$ (see their values in Table 2). The shape parameter value decreased consistently over the days. The scale parameter was high to begin with, then increased later. The distance, $d(y_1, \lambda)$ between the observable and non-observable Poisson mechanism for y_1 is calculated using the expression

$$d(y_1, \lambda) = \{\beta(1-\beta) \pm 1\}(\frac{\alpha}{\{1+\beta\}^2}) \quad (7)$$

and displayed in Table 2. Notice that the distance was large to begin with, then decreased but increased later over the days.

Note that we compute $\hat{p}_i = \frac{y_2}{y_1}$ for the i^{th} day. Then, we calculate the

$$\text{average:} \quad \bar{p} = \sum_{i=1}^{2} p_i / 2 \quad \text{and the variance:} \quad s_p^2 = \frac{(p_1 - p_2)^2}{4} \quad (\hat{p} \text{ in Table 1}),$$

and it had been steadily increasing over the days since 15th February 2020. This is something valuable for medical professionals learning the clinical nature of COVID-19. Using the expression,

$$odds_{Y_2} \approx e^{-p\lambda}\{1 + e^{-p(1-p)\lambda}\} \quad (8)$$

in Section 2.2, we calculated the odds for a COVID-19 case to become an asymptomatic type and displayed in Table 2.

Likewise, using the expression

$$odds_{Y_3} \approx e^{-(1-p)\lambda}\{1+e^{-p(1-p)\lambda}\}, \quad (9),$$

we estimated the odds for a COVID-19 case to become a symptomatic case as shown in Table 2. Notice that both odds ($Odds_{Y_2}$ and $Odds_{Y_3}$) are low but their odds ratio,

$$OR_{\frac{Y_3}{Y_2}} = e^{-(1-2p)\lambda} \quad (10)$$

is not negligible but reveals that the situation is favorable to symptomatic rather than asymptomatic. This discovery is feasible because of the approach, and it is an eye-opening reality for the medical professionals in their desire to control the spread of the COVID-19 virus. Both the observable, $\hat{H}_{y_2|y_1}$ and non-observable, $\hat{H}_{y_1,\gamma,\delta}$ binomial heterogeneity (see their values in Table 3) were decreasing for the number, y_2 of asymptomatic COVID-19 cases. The distance, $d(y_2, p)$ between the observable and non-observable for asymptomatic cases was moderate in the beginning, then increased, and then decreased over the next days (see their values in Table 3). However, the distance, $d(Y_2, Y_3)$ between the observable, y_2 of the asymptomatic cases and the observable, y_3 of the symptomatic cases was narrow, then wider, and then moderate over the days (their values in Table 3).

For a COVID-19 case to become a symptomatic type, the chance is moderate to less and then more over the days ($1-\bar{p}$ in Table 3). The estimate of the shape and scale parameter happened to be $\hat{\gamma}$ and $\hat{\delta}$ respectively (see their values in Table 3). Both the shape parameter and the scale parameter values decreased drastically over the days. From the p-values in Table 4, we infer that the prevalence rate, $\hat{\lambda}$, the distances, $d(y_1, \lambda)$, $d(y_2, p)$ and $d(Y_2, Y_3)$ do differ significantly over the three groups of dyad days. The chance for COVID-19 to become an asymptomatic type does not differ significantly across the three groups. On the contrary, the non-observable heterogeneities $H_{\hat{\lambda}}$ of the Poisson random number, y_1 and

257 $\hat{H}_{y_1,\gamma,\delta}$ of the binomial random number, y_2 are not significant. Likewise, the observable heterogeneities $\hat{H}_{y_1}$

258 of the Poisson random number, y_1 and $\hat{H}_{y_2|y_1}$ of the binomial random number, y_2 for a given y_1 are not

259 significant.

260

261 **Table 1. COVID-19 in Cruise Ship, 2020, Mizumoto et al. (2020)**

Date	Y_3	Y_2	Y_1	$\bar{\lambda} = \bar{y}_1$	$s_\lambda^2 = V\hat{a}r(Y_1)$	$OR_{\frac{1\to 2}{0\to 1}}$	$Odds_{Y_1}$
Feb 15-16, 2020	29, 32	38, 38	67, 70	68.5	4.5	0.5001	1.7E-30
Feb 17-18, 2020	29, 23	70, 65	99, 88	93.5	60.5	0.5000	2.4E-41
Feb 19-20, 2020	11, 7	68, 6	79, 13	46	21.78	0.5002	1.0E-20

262

263 **Table 2. Results for Mizumoto et al.'s COVID-19 Data in Diamond Princess**

Date	$OR_{\frac{Y_3}{Y_2}}$	$\hat{H}_{y_1}$	$\hat{\beta}$	$\hat{\alpha}$	$d(y_1,\lambda)$	$H_{\hat{\lambda}}$
15, 16 Feb 2020	943.88	**0.27**	15.22	1042.72	857.81	0.93
17, 18 Feb 2020	7.36E+17	**0.70**	1.54	144.50	18.79	0.61
19, 20 Feb 2020	9.69E+11	**0.65**	2.11	97.15	23.56	0.67

264

265 **Table 3. Results for Asymptomatic COVID-19 Cases in Mizumoto et al. (2020)**

Date	$1-\bar{p} = 1-Ave(\frac{y_2}{y_1})$	$s_p^2 = V\hat{a}r(\frac{y_2}{y_1})$	$\hat{H}_{y_1,\gamma,\delta}$	$\hat{H}_{y_2\|y_1}$	$d(y_2,p)$	$d(Y_2,Y_3)$
15, 16 Feb 2020	0.45	0.0002	**0.99**	**0.95**	37.125	6.85

17, 18 Feb 2020	0.28	0.0004	**0.98**	**0.89**	66.6	41.14
19, 20 Feb 2020	0.34	0.0796	0.10	**0.74**	29.7	14.72

Table 4. Tango's Test Statistic and Its P-Value for Several Entities

Tango statistic	$\hat{H}_{y_1}$	$H_{\hat{\lambda}}$	$\bar{p}$	$\hat{H}_{y_1,\gamma,\delta}$	$\hat{H}_{y_2\|y_1}$	$d(y_1,\lambda)$	$d(y_2,p)$	$d(Y_2,Y_3)$
T with 2 df	0.25	0.36	0.41	**0.77**	**0.51**	699420	260.66	751.20
p-value	0.87	0.83	0.81	**0.67**	**0.77**	0.0E100	2.4E-57	7.5E-164

5. DISCUSSION AND CONCLUSION

The risk of contracting the COVID-19 virus during a cruise is more than in a community setting, as confined spaces discourage non-pharmaceutical mitigation strategies such as social distancing to be weakly implemented and breathing air is tightly internalized. More nations are afraid to let the voyagers come ashore at the seaports. Ships are not even permitted to dock at the port, as to not complicate virus mitigation efforts by the local surrounding communities. The scenario seems to be anti-humanistic. The medical doctors and/or pharmaceutical service were strained due to the infected and COVID-19-free voyagers. Lack of clear symptoms among those that were infected added to difficulties in managing the COVID-19 crisis onboard the ship, and for any ship for that matter. Most importantly, how do we dispose of the COVID-19 fatalities (bodies), in a safe manner?

In the midst of uncertainties about the root cause and/or the appearance of any symptoms, the best modelers can do (as it is done in this article) is to devise a methodology to address the observable as well as non-observable heterogeneity, estimate the proportion of COVID-19 cases to be asymptomatic, estimate the odds of becoming symptomatic, and also the odds ratio for asymptomatic in comparison to those symptomatic among COVID-19 cases. Some of these are non-trivial to the professional experts

dealing with the intention of reducing the spread of COVID-19 if not its total control. Still much of COVID-19 is a mysterious pandemic. It is clear that non-pharmaceutical mitigation strategies such as social distancing, utilization of face coverings, frequent hand sanitization, infected people quarantining on board, and severely controlled ship cleanliness and sanitation standards are required; this may only be successful with limited numbers of passenger and crew members. Given the nature of the disease, its heterogeneity and human social norms, pre-voyage and post-voyage quick testing procedures may become the new standard for cruise ship passengers and crew. The technological advances in testing provided today would facilitate more humanistic treatment as compared to more archaic quarantine and isolation practices for all onboard ship. With quick testing, identification of those infected and thus not allowed to embark on a cruise or quarantine those disembarking, and other mitigation strategies, the popular cruise adventure could be available safely again. Whatever the procedures implemented, the methodological purpose of this study should add valuable insight in the modeling of disease and specifically, the COVID-19 virus.

Funding None.

Conflict of interest The authors have no conflict of interest.

Availability of data and materials There is no other data or materials other than what are in the manuscript itself.

Code availability None.

Authors' contribution The authors contributed everything in the manuscript.

Acknowledgements The authors thanks Texas State University and The University of Texas Health Science Center at Tyler for the support to write this article.

6. REFERENCES

1. Blumenfeld D. *Operations Research Calculations Handbook*, CRC Press, Boca Raton, Florida, 2020.
2. Ecochard J. Heterogeneity in fecundability studies: issues and modeling, *Statistical Methods in Medical Research; 2006, 15: 141-160.*
3. Elston R, Olson J, Palmer L. *Biostatistical Genetics and Genetic Epidemiology,* Wiley Press, Baffins Lane, Chichester, West Sussex, UK, 2002.
4. Hope T, Norris P. Heterogeneity in the frequency distribution of crime victimization, *Journal of Quantitative Criminology; 2013, 29 (4): 543-576.*
5. Hunink MG, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, Glasziou PF. *Decision making in health and medicine: Integrating evidence and values,* Cambridge University Press, Cambridge, UK, 2018.
6. Khokhlov V. Conditional value-at-risk for elliptical distributions, *Evropský časopis Ekonomiky a Management*; 2016, 2 (6): 70–79.
7. Mizumoto K, Chowell G. Transmission potential of the novel coronavirus (COVID-19) onboard the Diamond Princess Cruises Ship, 2020, *Infectious Disease Modelling; 2020, 5: 264-270.*
8. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19), cases on board the Diamond Princess cruise ship, Yokohama, 2020. *Euro Surveill;2020, 25(10): 1560-7917.*
9. Rajan C, Shanmugam R. *Discrete Distributions in Engineering and the Applied Sciences, Synthesis Lectures on Mathematics and Statistics, Vol. 12, Pages 1-227*, Morgan & Claypool Press, 82 Winter Sport Ln, Williston, VT 05495, USA, 2020.

10. Ross S. (2002) *A First Course in Probability*, sixth edition, Prentice Hall, Upper Saddle River, New Jersey 07458, 2002.
11. Shanmugam R. Probabilistic Patterns among Coronavirus Confirmed, Cured and Deaths in Thirty-two India's States/Territories, *International Journal of Ecological Economics and Statistics, 41 (4), 45-56*, 2020, in press.
12. Shanmugam R, Radhakrishnan R. Incidence jump rate reveals over/under dispersion in count data, *International Journal of Data Analysis, and Information Systems*, 3(1), 1-8, 2011.
13. Spreeuw J. Heterogeneity in hazard rates in insurance, Tinbergen Institute of Research Series, 210, *Ph. D. thesis*, University of Amsterdam, Amsterdam, 1999.
14. Stuart A, Ord K. *Kendall's Advanced Theory of Statistics, Volume 1*. Oxford University Press, London, 2015.
15. Tango T. The detection of disease clustering in time, *Biometrics; 1984, 40: 15-26.*

APPENDIX I

Poisson Heterogeneity: Derivations

It is known that the conjugate prior for the Poisson distribution is gamma, whose pdf is

$$c(\lambda|\alpha,\beta)d\lambda = e^{-(\alpha\lambda)}(\alpha\lambda)^{\beta-1}d(\alpha\lambda)/\Gamma(\beta);\alpha>0;\beta>0\,, \tag{1}$$

with an average. $E(\lambda|\alpha,\beta)=\frac{\beta}{\alpha}$ and variability $Var(\lambda|\alpha,\beta)=E(\lambda|\alpha,\beta)/\alpha$, where the parameters α and β are recognized as *hyper-parameters* (Rajan and Shanmugam, 2020). Notice that the hyper parameter $\alpha>0$ causes the variability in the COVID-19's prevalence rate to fluctuate up or down and hence, you would anticipate the heterogeneity to involve the hyperparameter α.

$$c(\lambda|y_1,\alpha,\beta)=\Pr(y_1|\lambda)c(\lambda|\alpha,\beta)/\int_{-\infty}^{\infty}\Pr(y_1|\lambda)c(\lambda|\alpha,\beta)d\lambda \tag{2}$$

is the posterior pdf of the non-observable λ. Also, the denominator

$$\int_0^{\infty}\Pr(y_1|\lambda)c(\lambda|\alpha,\beta)d\lambda=\Gamma(\alpha+y_1)/(1+\beta)^{\alpha+y_1}\,,$$

in a Bayesian framework, is called the *marginal distribution*. With $\Delta_\lambda=\lambda-E(\lambda)$, it is clear that $\int_{-\infty}^{\infty}\Delta_\lambda c(\lambda|\alpha,\beta)d\lambda=0$, note that the prior variance is

$$Var(\lambda|\alpha,\beta)=\int_{-\infty}^{\infty}\Delta_\lambda^2 c(\lambda|\alpha,\beta)d\lambda\,.$$

Because the prior is conjugate, its counterpart's variability

$$Var(\lambda|y_1,\alpha,\beta)=\int_{-\infty}^{\infty}(\lambda-E[\lambda|y_1,\alpha,\beta)^2 c(\lambda|\alpha,\beta)d\lambda$$

is minimal when the Bayes estimate of the non-observable is the posterior mean, $\lambda_{Bayes} = E[\lambda | y_1 \alpha, \beta]$, where

$$E[\lambda | y_1, \alpha, \beta] = \frac{(\alpha + \overline{y}_1)}{\beta}.$$

Differentiating the log-likelihood function

$$\ln L(n\overline{y}_1, \lambda) = n\overline{y}_1 \ln \lambda - n\lambda + \sum_{i=1}^{n} \ln(y_i!)$$

with respect to the non-observable parameter, λ, setting it equal to zero and solving it, we obtain the MLE and it is $\hat{\lambda}_{mle} = \overline{y}_1$. Because of the invariance property of the MLE, it is involved. The invariance property refers to that the MLE of a function of the parameter is the function of the MLE of the parameter. Also, it is known (Blumenfeld, 2010) that

$$E_{prior} E_{ikelihood}(\overline{y}_1 | \lambda) = E(\overline{y}_1) \quad \text{and}$$

$$Var(\overline{y}_1) = E_{prior} Var_{likelihood}(\overline{y}_1 | \lambda) + Var_{prior} E_{likelihood}(\overline{y}_1 | \lambda). \qquad (3)$$

Hence, we are ready now to define the non-observable heterogeneity below in the Definition 1.

Definition 1. The non-observable heterogeneity of the Poisson parameter, λ is defined as

$$H_\lambda = [1 + \frac{Var_{prior} E_{likelihood}(\overline{y}_1 | \lambda)}{E_{prior} Var_{likelihood}(\overline{y}_1 | \lambda)}]^{-1} \in [0,1]. \qquad (4)$$

Following the Definition 1, we obtain the *non-observable heterogeneity* of the COVID-19 cases is

389 $$H_{\lambda} = [1 + \frac{1}{\beta}]^{-1} \in [0,1] \quad (5)$$

390 When the value of H_{λ} is closer to zero, the data are believed to have non-observable Poisson

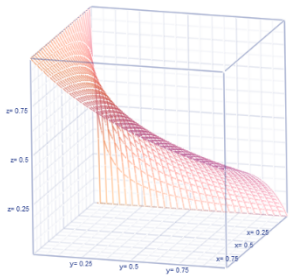
391 homogeneity. Its MLE is

392 $$\hat{H}_{\lambda} = [1 + \frac{1}{\hat{\beta}}]^{-1} = [1 + \frac{s_{\lambda}^2}{\overline{\lambda}}]^{-1} = \frac{\overline{\lambda}}{\overline{\lambda} + s_{\lambda}^2}. \quad (6)$$

393

394 The reader is referred to Figure 1 for the configuration of the non-observable Poisson heterogeneity in

395 general.



396

397 **Figure 1. Non-observable Heterogeneity**

398

399 Likewise, the *observable-heterogeneity* is defined below in Definition 2.

400

401 *Definition2.* The observable heterogeneity of the randomly sampled Poisson counts, $y_1, y_2, ..., y_n$ is

402 defined as

403 $$H_{\overline{y}_1} = [1 + \frac{Var_{marginal} E_{posterior}(\lambda | \overline{y}_1)}{E_{marginal} Var_{posterior}(\lambda | \overline{y}_1)}]^{-1} \in [0,1]. \quad (7)$$

404

Before we apply the Definition 2, let us recollect that the marginal pdf of the complete sufficient statistic, $\overline{y}_1$ is uniform distribution and the posterior distribution is

$$c(\lambda \mid \overline{y}_1, \alpha, \beta) = (1+\beta)^{(\alpha + n\overline{y}_1)} [e^{-(1+\beta)\lambda}]^{(\alpha + n\overline{y}_1)-1} / \Gamma(\alpha + n\overline{y}_1)$$

with

$$E(\lambda \mid \overline{y}_1, \alpha, \beta) = \frac{(\alpha + n\overline{y}_1)}{(1+\beta)} \tag{8}$$

and

$$Var(\lambda \mid \overline{y}_1, \alpha, \beta) = \frac{E(\lambda \mid \overline{y}_1, \alpha, \beta)}{(1+\beta)}. \tag{9}$$

Imposing the Definition 2 and simplifying, we obtain that $H_{\overline{y}_1} = [1 + \frac{(1+\beta)}{6}]^{-1}$ whose MLE is

$$\hat{H}_{\overline{y}_1} = [1 + \frac{(1+\hat{\beta})}{6}]^{-1} = [1 + \frac{(1+\frac{\overline{\lambda}}{s_\lambda^2})}{6}]^{-1} \in [0,1]. \tag{10}$$

The reader is referred to Figure 2 for the configuration of the observable Poisson heterogeneity, $\hat{H}_{\overline{y}_1}$ in general. When the value of $\hat{H}_{\overline{y}_1}$ is closer to zero, the data are interpreted to have observable homogeneity.

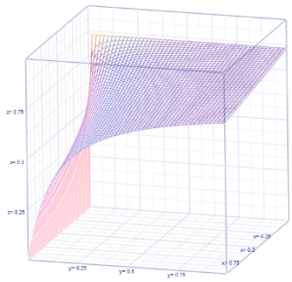


Figure 2. Observable Heterogeneity

Furthermore, the distance, $d(y_1,\lambda)$ between the observable y_1 of the number of COVID-19 cases and the prevalence rate λ could be assessed using the formula

$$d(y_1,\lambda) = E_{Y_1}E_{\lambda}\left|Y_1-\lambda\right| = \sum_{y_{1=0}}^{\infty}\int_{0}^{\infty}\left|Y_1-\lambda\right|\Pr(y_1\left|\lambda\right.)c(\lambda\left|\overline{y}_1,\alpha,\beta\right.)d\lambda\,. \tag{11}$$

Realizing that their absolute difference is really $\left|Y_1-\lambda\right| = Y_1+\lambda-2\min\{Y_1,\lambda\}$, we obtain after simplifications that

$$d(y_1,\lambda) = \{\beta(1-\beta)\pm 1\}\left(\frac{\alpha}{\{1+\beta\}^2}\right). \tag{12}$$

The configuration of the distance, $d(y_1,\lambda)$ between the observable and non-observable in Poisson mechanism. We now turn to discuss stochastic properties of the Poisson distribution are given in Figure 3.

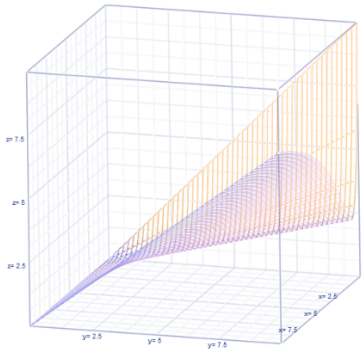


Figure 3. Distance, $d(y_1,\lambda)$ in Poisson.

The survival function of the random number, Y_1 of COVID-19 cases is

$$S_{Y_1}(r\left|\lambda\right.) = \Pr(Y_1\geq r\left|\lambda\right.) = \sum_{i=r}^{\infty}e^{-\lambda}\lambda^i/i! = P[\chi^2_{2(r+1)df} < 2\lambda]; \lambda > 0. \tag{13}$$

The hazard rate is a force of mortality. The hazard rate, $h(y_1)$ for the COVID-19 occurrence is

$$h(y_1) = \frac{\Pr(y_1|\lambda)}{S(y_1+1|\lambda)} = \frac{e^{-\lambda}\lambda^{y_1}}{y_1!P[\chi^2_{2(y_1+2)df} < 2\lambda]}; \lambda > 0. \quad (14)$$

Does the Poisson chance mechanism keep any a finite *memory?* For example, the geometric distribution is known to have no memory. What is memory? The memory is really a conditional probability. That is,

$$memory = \Pr(Y_1 \geq s | y_1 \geq r) = \frac{\Pr(Y_1 \geq r+s)}{\Pr(Y_1 \geq r)} = \frac{P[\chi^2_{2(r+s+1)df} < 2\lambda]}{P[\chi^2_{2(r+1)df} < 2\lambda]}; \lambda > 0, \quad (15)$$

confirming that there is a finite memory in the Poisson mechanism of COVID-19 incidences. To be specific, with $r=0, s=1$ in the above result, the memory between COVID-19 free situation and just one COVID-19 occurrence is revealed in the chance-oriented Poisson mechanism. Such a memory is

$$memory_{0\to1} = \frac{P[\chi^2_{4df} < 2\lambda]}{P[\chi^2_{2df} < 2\lambda]}; \lambda > 0. \quad (16)$$

Likewise, the memory between at least one COVID-19 case situation and at least two COVID-19 cases situation is revealed with a substitution of $r=1, s=1$ in the above result and it is

$$memory_{1\to2} = \frac{P[\chi^2_{6df} < 2\lambda]}{P[\chi^2_{4df} < 2\lambda]}; \lambda > 0. \quad (17)$$

The odds ratio from the initial $memory_{0\to1}$ to the next $memory_{1\to2}$ is

$$OR_{\frac{1\to2}{0\to1}} = \frac{P[\chi^2_{6df} < 2\lambda]P[\chi^2_{2df} < 2\lambda]}{\{P[\chi^2_{4df} < 2\lambda]\}^2} \quad (18)$$

454

455 (their values in Table 1). However, the odds for COVID-19 free healthy situation to prevail is

456

457 $$Odds_{Y_1} = \frac{\Pr(Y_1 = 0)}{\Pr(Y_1 \geq 1)} = (e^{\lambda} - 1)^{-1}; \lambda > 0 \qquad (19)$$

458

459 (their values in Table 1). For details on how the chance for an incidence of a disease to occur from a

460 disease-free scenario changes, the reader is referred to Shanmugam and Radhakrishnan (2011).

461

APPENDIX II

Binomial Heterogeneity: Derivations

Let an indicator random variable, $I_i = 1$ for a COVID-19 case to be asymptomatic with a probability, $0 < p < 1$ and $I_i = 0$ for the case to be symptomatic with a probability, $0 < 1 - p < 1$. Then, for a fixed y_1, the random variable, $Y_2 = \sum_{i=1}^{Y_1} I_i$ follows a binomial probability distribution with parameters (y_1, p). Likewise, for a fixed y_1, the random variable, $Y_3 = y_1 - Y_2$ follows a complementary binomial distribution with parameters $(y_1, 1-p)$. That is,

$$\Pr(Y_2 = y_2 \,|\, y_1, p) = \binom{y_1}{y_2} p^{y_2} (1-p)^{y_1 - y_2}; y_2 = 0, 1, 2, ..., y_1; 0 < p < 1 \quad (20)$$

and

$$\Pr(Y_3 = y_3 \,|\, y_1, p) = \binom{y_1}{y_3} (1-p)^{y_3} p^{y_1 - y_3}; y_3 = 0, 1, 2, ..., y_1; 0 < 1 - p < 1 \quad (21)$$

with their conditional expected numbers

$$E(Y_2 \,|\, y_1, p) = y_1 p \;\; E(Y_3 \,|\, y_1, 1-p) = y_1 (1-p) = y_1 - E(Y_2 \,|\, y_1, p) \quad (22)$$

and the conditional variabilities

$$Var(Y_2 \,|\, y_1, p) = (1-p) E(Y_2 \,|\, y_1, p), \quad (23)$$

and

$$Var(Y_3 \,|\, y_1, 1-p) = p E(Y_3 \,|\, y_1, 1-p). \quad (24)$$

The conditional variability of Y_2 is a percent $(1-p)$ of its expected number $E(Y_2 \,|\, y_1, p)$, implying that it exhibits under dispersion. Likewise, the conditional variability of Y_3 is a percent $(1-p)$ of its

481 expected number $E(Y_3 \mid y_1, p) = y_1(1-p)$ implying that it also exhibits under dispersion. Together, the

482 above statements suggest a conditional balance

483 $$\frac{E(Y_2 \mid y_1, p)}{E(Y_3 \mid y_1, 1-p)} = odds(asmptomatic) = \frac{p}{(1-p)} \tag{25}$$

484 (Stuart and Ord, 2015 for details of the odds concepts). Consequently, we note that

485 $$p = \frac{E(Y_2 \mid y_1, p)}{E(Y_2 \mid y_1, p) + E(Y_3 \mid y_1, 1-p)}. \tag{26}$$

486 Furthermore, we wonder whether the random variables Y_2 and Y_3 are correlated? The answer is

487 affirmative. To identify their correlation, notice that

488 $$E(Y_2) = E_{Y_1} E(Y_2 \mid y_1) = E_{Y_1}(Y_1 p) = p\lambda,$$

489 $$E(Y_3) = E_{Y_1} E(Y_3 \mid y_1) = E_{Y_1}(Y_1\{1-p\}) = (1-p)\lambda$$

490 $$Var(Y_2) = E_{Y_1} Var(Y_2 \mid y_1) + Var_{Y_1} E(Y_2 \mid y_1) = E_{Y_1}\{Y_1 p(1-p)\} + Var_{Y_1}(Y_1 p) = p\lambda$$

491 $$Var(Y_3) = E_{Y_1} Var(Y_3 \mid y_1) + Var_{Y_1} E(Y_3 \mid y_1) = E_{Y_1}\{Y_1 p(1-p)\} + Var_{Y_1}(Y_1\{1-p\}) = (1-p)\lambda$$

492 $$Cov(Y_2, Y_3) = E_{Y_1} E(Y_2 Y_3 \mid y_1) - E_{Y_1} E(Y_2 \mid y_1) E_{Y_1} E(Y_3 \mid y_1)$$

493 where

494 $$E_{Y_1} E(Y_2 \mid y_1) = p\lambda, E_{Y_1} E(Y_3 \mid y_1) = (1-p)\lambda,$$

495 $$E_{Y_1} E(Y_2 Y_3 \mid y_1) = E_{Y_1} E_{Y_2} E_{Y_3 \mid Y_2, Y_1}(Y_2 Y_3 \mid Y_1) = E_{Y_1} E_{Y_2 \mid Y_1}\{Y_2 Y_1(1-p)\} = E_{Y_1}\{p(1-p)Y_1^2\} = p(1-p)\lambda(1+\lambda).$$

496

497 Hence, their correlation is

498 $$\rho_{Y_2, Y_3} = \frac{Cov(Y_2, Y_3)}{\sqrt{Var(Y_2) Var(Y_3)}} = \sqrt{p(1-p)}. \tag{27}$$

Their expected distance, $d(Y_2, Y_3) = E_{Y_1} E(|Y_2 - Y_3||Y_1)$ portrays the drift between the symptomatic observable, Y_2 and the asymptomatic observable, Y_3 and it is simplified to this function $d(Y_2, Y_3) = |2p - 1|\lambda$ (see Table 3 for their values), due to applying

$$|Y_2 - Y_3| = Y_2 + Y_3 - 2\min\{Y_2, Y_3\}.$$

Let us assume that every COVID-19 case has the same chance of being asymptomatic in a time period. Then, the random number, y_2 for a specified number, y_1 of COVID-19 cases follows a binomial distribution with parameters (y_1, p). We select a conjugate beta prior distribution

$$c(p|\gamma, \delta)dp = \Gamma(\gamma + \delta) p^{\gamma - 1}(1 - p)^{\delta - 1} / \Gamma(\gamma)\Gamma(\delta); 0 < p < 1; \gamma, \delta > 0 \quad (28)$$

for our discussion for asymptomatic COVID-19 cases. The prior average is

$$\mu_{prior} = E(p|\gamma, \delta) = \frac{\gamma}{(\gamma + \delta)}$$

and the prior variability is

$$Var(p|\gamma, \delta) = \mu_{prior}(1 - \mu_{prior}) / (1 + \gamma + \delta),$$

where the parameters γ and δ are hyper-parameters (Rajan and Shanmugam, 2020, for details). We guess that the binomial heterogeneity would involve both hyper parameters. The task for us is how do we construct such heterogeneity? An answer is the following. The posterior distribution

$$\begin{aligned} c(p|\overline{y}_1, \overline{y}_2, \gamma, \delta) &= \Pr(\overline{y}_2|\overline{y}_1, p) c(p|\gamma, \delta) / \int_{-\infty}^{\infty} \Pr(\overline{y}_2|\overline{y}_1, p) c(p|\gamma, \delta) dp \\ &= p^{\gamma + \overline{y}_2 - 1}(1 - p)^{\delta + \overline{y}_1 - \overline{y}_2 - 1} / \{\Gamma(\gamma + \overline{y}_2)\Gamma(\delta + \overline{y}_1 - \overline{y}_2) / \Gamma(\gamma + \delta + \overline{y}_1)\} \end{aligned} \quad (29)$$

519 would play a key role to construct both the observable and non-observable binomial heterogeneity. With

520 $\Delta_p = p - E(p)$, it is clear that $\int_{-\infty}^{\infty} \Delta_\lambda c(p|\gamma,\delta)dp = 0$.

521 The prior variance is

522 $$Var(p|\gamma,\delta) = \int_{-\infty}^{\infty} \Delta_\lambda^2 c(p|\gamma,\delta)dp .$$

523 Its posterior counterpart

524 $$Var(p|\overline{y}_1,\overline{y}_2,\gamma,\delta) = \int_{-\infty}^{\infty} (p - E[p|\overline{y}_1,\overline{y}_2,\gamma,\delta)^2 c(p|\gamma,\delta)dp$$

525 is minimal when the Bayes estimate of non-observable is the posterior mean

526 $p_{Bayes} = E(p|\overline{y}_1,\overline{y}_2,\gamma,\delta)$,

527 where

528 $$\mu_{posterior} = E(p|\overline{y}_1,\overline{y}_2,\gamma,\delta) = \frac{(\gamma+\overline{y}_2)}{(\gamma+\delta+\overline{y}_1)}. \quad (30)$$

529 The posterior variance is

530 $$Var(p|\overline{y}_1,\overline{y}_2,\gamma,\delta) = \frac{\mu_{posterior}(1-\mu_{posterior})}{(1+\gamma+\delta+\overline{y}_1)}. \quad (31)$$

531 Differentiating the log-likelihood function as

532 $$\ln L(n,\overline{y}_1,\overline{y}_2,p) = \overline{y}_2 \ln p + (\overline{y}_1 - \overline{y}_2)\ln(1-p) + \sum_{i=1}^{n} \ln\binom{y_i}{y_{2,i}}$$

533 with respect to the non-observable parameter, p, setting it equal to zero and solving it, we obtain the

534 MLE and it is $\hat{p}_{mle} = \frac{\overline{y}_2}{\overline{y}_1}$ It is known that

535 $$E_{prior}E_{likelihood}(\overline{y}_2|\overline{y}_1,p) = E(\overline{y}_2) \quad (32)$$

and

$$Var(\overline{y}_2) = E_{prior} Var_{likelihood}(\overline{y}_2 \,|\overline{y}_1, p) + Var_{prior} E_{likelihood}(\overline{y}_2 \,|\overline{y}_1, p)\,. \tag{33}$$

Hence, we define the non-observable binomial heterogeneity below in Definition 3.

Definition 3. The non-observable binomial heterogeneity is defined as

$$H_{y_1,p} = [1 + \frac{Var_{prior} E_{likelihood}(\overline{y}_2 \,|\overline{y}_1, p)}{E_{prior} Var_{likelihood}(\overline{y}_2 \,|\overline{y}_1, p)}]^{-1} \in [0,1]. \tag{34}$$

Following the Definition 3, we obtain the *non-observable heterogeneity* of the COVID-19's asymptotic cases (remembering that $(\overline{y}_1, \gamma, \delta)$ are the non-observable parameters) as

$$H_{y_1,\gamma,\delta} = [1 + y_1 \frac{Var_{prior}(p)}{E_{prior}(p\{1-p\})}]^{-1} = [1 + \frac{y_1}{(\gamma+\delta)(1+\gamma+\delta)}]^{-1} \in [0,1]. \tag{35}$$

When the value of $H_{y_1,\gamma,\delta}$ is closer to zero, the data are interpreted to have non-observable binomial homogeneity. Substituting the MLEs

$$\hat{\gamma} = \overline{p}\{\frac{\overline{p}(1-\overline{p})}{s_p^2} - 1\} \text{ and } \hat{\delta} = \frac{(1-\overline{p})\hat{\gamma}}{\overline{p}}, \tag{36}$$

we obtain its MLE

$$\hat{H}_{y_1,\gamma,\delta} = [1 + \frac{y_1 (s_p^2)^2}{(\left|\overline{p}(1-\overline{p}) - s_p^2\right|)\overline{p}(1-\overline{p})}]^{-1} \in [0,1]. \tag{37}$$

Likewise, the *observable-heterogeneity* of the binomial distribution of y_2 is defined below in Definition 4.

555 *Definition 4.* The observable heterogeneity of the binomial counts, $y_{2,i}, i = 1, 2, ..., y_1$ (in terms of the

556 complete sufficient statistic $\overline{y}_2$) is defined as

557 $$H_{y_2} = [1 + \frac{Var_{marginal} E_{posterior}(p|\overline{y}_1)}{E_{marginal} Var_{posterior}(p|\overline{y}_1)}]^{-1} \in [0,1]. \quad (38)$$

558 Before we apply Definition 4, remember that the marginal pdf of the complete sufficient statistic, $\overline{y}_2$ is

559 the beta-binomial distribution,

560 $$\Pr(\overline{Y}_2) = \binom{\overline{y}_1}{\overline{y}_2} \frac{\Gamma(\gamma + \overline{y}_2)\Gamma(\delta + \overline{y}_1 - \overline{y}_2)}{\Gamma([\gamma + \delta] + \overline{y}_1)\}}; \quad (39)$$

561 and the posterior distribution is beta. With the notation $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, we note that the probability

562 mass function of the beta-binomial distribution is

563 $$\Pr(y_2) = \binom{y_1}{y_2} B(\gamma + y_2, \delta + y_1 - y_2) / B(\gamma, \delta); y_2 = 0, 1, 2..., y_1; \gamma, \delta > 0. \quad (40)$$

564 That is, the posterior probability density function is

565 $$c(p|\overline{y}_1, \overline{y}_2, \gamma, \delta) = \frac{\Gamma(\gamma + \delta + \overline{y}_1)\}}{\Gamma(\gamma + \overline{y}_2)\Gamma(\delta + \overline{y}_1 - \overline{y}_2)} p^{\gamma + \overline{y}_2 - 1}(1-p)^{\delta + \overline{y}_1 - \overline{y}_2 - 1} \quad (41)$$

566 with

567 $$E(p|\overline{y}_1, \overline{y}_2, \gamma, \delta) = \frac{(\gamma + \overline{y}_2)}{([\gamma + \delta] + \overline{y}_1))} \quad (42)$$

568 and

569 $$Var(p|\overline{y}_1, \overline{y}_2, \gamma, \delta) = \frac{(\gamma + \overline{y}_2)(\delta + \overline{y}_1 - \overline{y}_2)}{([\gamma + \delta] + \overline{y}_1)(1 + [\gamma + \delta] + \overline{y}_1)}. \quad (43)$$

570 Now applying Definition 4, we obtain an expression for the observable binomial heterogeneity

$$H_{y_2|y_1} = [1 + \frac{Var_{marginal}\{\frac{(\gamma + y_2)}{(\gamma + \delta + y_1)}\}}{E_{marginal}\{\frac{(\gamma + y_2)(\delta + y_1 - y_2)}{(\gamma + \delta + y_1)(1 + \gamma + \delta + y_1)}\}}]^{-1} \approx [1 + (\frac{\delta}{\delta + \gamma})(\frac{y_1}{y_1 + \delta})]^{-1} \in [0,1], \quad (44)$$

whose estimate is

$$H_{y_2|y_1} \approx [1 + \overline{p}(\frac{y_1 s_p^2}{y_1 s_p^2 + \{1 - \overline{p}\}\left|\overline{p}(1 - \overline{p}) - s_p^2\right|})]^{-1}, \quad (45)$$

Because

$$\hat{\gamma} \approx \overline{p}\{\frac{\overline{p}(1 - \overline{p})}{s_p^2} - 1\} \quad \text{and} \quad \hat{\delta} \approx \frac{\hat{\gamma}(1 - \overline{p})}{\overline{p}}. \quad (46)$$

When the value of $\hat{H}_{y_2|y_1}$ is closer to zero, the data are considered to have observable binomial homogeneity. Also, the distance, $d(y_2, p)$ between the observable y_2 of the number of asymptomatic COVID-19 cases and its proportion, p could be assessed using the formula

$$d(y_2, p) = E_{Y_2} E_p \left|Y_2 - p\right| = \sum_{y_2=0}^{y_1} \int_0^{\infty} \left|Y_2 - p\right| \Pr(y_2 \left| p) c(p \right| y_1, y_2, \gamma, \delta) dp. \quad (47)$$

Realizing that the absolute difference, $\left|Y_2 - p\right| = Y_2 + p - 2\min\{Y_2, p\}$, we obtain after simplifications that

$$d(y_2, p) = \left|y_1 - 1\right|(\frac{\gamma}{\gamma + \delta}). \quad (48)$$

Likewise, to obtain the *non-observable heterogeneity* of the COVID-19's symptomatic cases, all we have to do is change p to $(1 - p)$, change y_2 to y_3, along with changing γ to δ and go through the process above. Hence, the non-observable heterogeneity in the symptomatic cases is the same. That is,

$$H_{y_1,\delta,\gamma} = [1 + y_1 \frac{Var_{prior}(1 - p)}{E_{prior}(p\{1 - p\})}]^{-1} = [1 + \frac{y_1}{(\gamma + \delta)(1 + \gamma + \delta)}]^{-1} \in [0,1]. \quad (49)$$

The observable binomial heterogeneity for the symptomatic cases is

$$H_{y_3|y_1} \approx [1+(\frac{\gamma}{\delta+\gamma})(\frac{y_1}{y_1+\gamma})]^{-1} \in [0,1], \tag{50}$$

whose MLE is

$$\hat{H}_{y_3|y_1} \approx [1+\overline{p}(\frac{y_1 s_p^2}{y_1 s_p^2 + \overline{p}\left|\overline{p}(1-\overline{p})-s_p^2\right|})]^{-1}, \tag{51}$$

which is interestingly not the same as $\hat{H}_{y_2|y_1}$. Also, the distance, $d(y_3,1-p)$ between the observable y_3 of the number of asymptomatic COVID-19's symptomatic cases and the proportion, $1-p$ could be assessed using the formula

$$d(y_3,1-p) = E_{Y_2}E_p\left|Y_2-(1-p)\right| = \sum_{y_{3=0}}^{y_1}\int_0^{\infty}\left|Y_3-(1-p)\right|\Pr(y_3\left|1-p\right.)c(1-p\left|y_1,y_2,\gamma,\delta\right.)d(1-p) \tag{52}$$

and it is after simplifications that

$$d(y_3,p) = \left|y_1-1\right|(\frac{\delta}{\gamma+\delta}). \tag{53}$$

Now we explore statistical properties of the asymptomatic cases, y_2. The survival function of the random number, Y_2 with asymptotic symptoms is

$$S_{Y_2}(r,p\left|y_1\right.) = \Pr(Y_2 \geq r\left|y_1\right.) = \sum_{i=r}^{\infty}\frac{y_1!}{i!(y_1-i)!}p^i(1-p)^{y_1-i} = P[F_{(2r,2[y_1-r+1])df} \leq \frac{y_1 p(y_1-r+1)}{(1-p)r}]; 0<p<1. \tag{54}$$

The hazard rate, $h(y)$ of the binomial distribution for the asymptomatic cases is

$$h(y_2) = \frac{\Pr(y_2\left|p\right.)}{S(y_2+1\left|p\right.)} = \frac{y_1!\{\frac{p}{1-p}\}^{y_2}\{1-p\}^{y_1}}{y_2!(y_1-y_2)!P[F_{(2r,2[y_1-r+1])df} \leq \frac{y_1 p(y_1-r+1)}{\{1-p\}r}]}; 0<p<1. \tag{55}$$

The binomial distribution has a finite *memory*

$$\Pr(Y_2 \geq s \,|\, y_2 \geq r) = \frac{\Pr(Y_2 \geq r+s)}{\Pr(Y_2 \geq r)} = \frac{P[F_{(2r,2[y_1-(r+s)+1])df} \leq \dfrac{y_1 p(y_1-(r+s)+1)}{\{1-p\}(r+s)}]}{P[F_{(2r,2[y_1-r+1])df} \leq \dfrac{y_1 p(y_1-r+1)}{\{1-p\}r}]}. \quad (56)$$

confirming that the usual binomial distribution does possess a finite memory. The conditional odds, for a fixed y_1, for *safe* asymptomatic symptom are

$$Odds_{Y_2|y_1} = \frac{\Pr(Y_2=0)}{\Pr(Y_2 \geq 1)} = (1-p)^{y_1}\{1-(1-p)^{y_1}\}^{-1} \approx (1-p)^{y_1}\{1+(1-p)^{y_1}\}. \quad (57)$$

The unconditional odds for safe asymptotic symptom are

$$odds_{Y_2} \approx \sum_{y_1=0}^{\infty} Odds_{Y_2|y_1} \Pr[Y_1 = y_1 \,|\, \lambda] \approx \sum_{y_1=0}^{\infty} (1-p)^{y_1}\{1+(1-p)^{y_1}\}e^{-\lambda}\lambda^{y_1}/y_1! \approx e^{-p\lambda}\{1+e^{-p(1-p)\lambda}\} \quad (58)$$

The reader is referred to Figure 4 for the configuration of the odds in asymptotic COVID-19 occurrences in general.

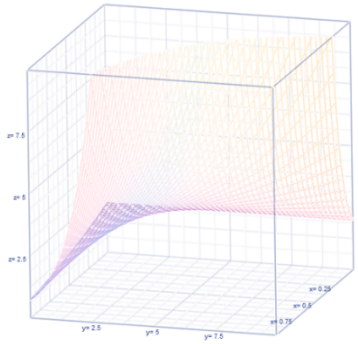


Figure 4. Odds for Asymptotic

Recall that $S_{Y_2}(1, p \,|\, y_1) = \Pr(Y_2 \geq 1 \,|\, y_1)$ is the likelihood for the existence of asymptomatic presentation of COVID-19 in the ship. The hazard in that situation (that is, with $r = 1$) is

$$h_{Y_2}(1|y_1,p)=1-\frac{P[F_{(4,2[y_1-1])df}\le\frac{y_1\hat{p}_{mle}(y_1-1)}{2(1-\hat{p}_{mle})}]}{P[F_{(4,2y_1)df}\le\frac{{y_1}^2\hat{p}_{mle}}{(1-\hat{p}_{mle})}]}, \tag{59}$$

where $\hat{p}_{mle}=\bar{y}_2$. A popular statistical concept in the business world (Khokhlov, 2016 for details), *Tail Value at Risk* (Tar) is

$$TVaR_{Y_2}=E[Y_2|Y_2\ge 1,p,y_1]\approx 1+\frac{y_1(1-p)}{p^2P[F_{(2r,2y_1)df}\le\frac{{y_1}^2p}{(1-p)}]}. \tag{60}$$

Similarly, all the Bayesian results for the binomial random variable, y_3 are easily derivable by interchanging γ and δ in all above expressions. The survival function of the random number, Y_3 with symptomatic symptoms is

$$S_{Y_3}(r,p|y_1)=\Pr(Y_3\ge r|y_1)=\sum_{i=r}^{\infty}\frac{y_1!}{i!(y_1-i)!}(1-p)^i\,p^{y_1-i}=P[F_{(2r,2[y_1-r+1])df}\le\frac{y_1(1-p)(y_1-r+1)}{pr}];0<p<1. \tag{61}$$

The hazard rate, $h(y)$ for the symptomatic sign is

$$h(y_3)=\frac{\Pr(y_3|p)}{S(y_3+1|p)}=\frac{y_1!\{\frac{1-p}{p}\}^{y_3}\,p^{y_1}}{y_3!(y_1-y_2)!P[F_{(2r,2[y_1-r+1])df}\le\frac{y_1(1-p)(y_1-r+1)}{pr}]};0<p<1. \tag{62}$$

The binomial distribution of those with symptomatic signs has a finite *memory*

$$\Pr(Y_3\ge s|y_3\ge r)=\frac{\Pr(Y_3\ge r+s)}{\Pr(Y_3\ge r)}=\frac{P[F_{(2r,2[y_1-(r+s)+1])df}\le\frac{y_1(1-p)(y_1-(r+s)+1)}{p(r+s)}]}{P[F_{(2r,2[y_1-r+1])df}\le\frac{y_1(1-p)(y_1-r+1)}{pr}]},$$

634 (63)

635 confirming that the usual binomial probability trend of those with symptomatic signs does possess a

636 finite memory. The conditional odds, for a fixed y_1, for *safe* symptomatic symptom are

637 $$Odds_{Y_3|y_1} = \frac{\Pr(Y_3 = 0)}{\Pr(Y_3 \geq 1)} = p^{y_1}\{1 - p^{y_1}\}^{-1} \approx p^{y_1}\{1 + p^{y_1}\}. \quad (64)$$

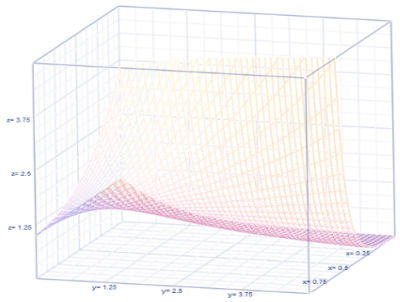
638 The unconditional odds for safe symptomatic symptom are

639 $$odds_{Y_3} \approx \sum_{y_1=0}^{\infty} Odds_{y_1} \Pr[Y_1 = y_1 | \lambda] \approx \sum_{y_1=0}^{\infty} p^{y_1}\{1 + p^{y_1}\} e^{-\lambda} \lambda^{y_1} / y_1! \approx e^{-(1-p)\lambda}\{1 + e^{-p(1-p)\lambda}\}. \quad (65)$$

640 A comparison of $odds_{Y_2}$ and $odds_{Y_3}$ suggests the odds ratio,

641 $$OR_{Y_3/Y_2} = \frac{odds_{Y_3}}{odds_{Y_2}} = e^{-(1-2p)\lambda}. \quad (66)$$

642 See Figure 5 for the configuration of the isomorphic factor, $e^{-(1-2p)\lambda}$.



643

644 **Figure 5. The configuration isomorphic factor** $e^{-(1-2p)\lambda}$

645

646 Recall that $S_{Y_3}(1, p | y_1) = \Pr(Y_3 \geq 1 | y_1)$ is the chance for the existence of symptomatic symptom of

647 COVID-19. The hazard in that situation (that is, with $r = 1$) is

$$h_{Y_3}(1|y_1,p) = 1 - \frac{P[F_{(4,2[y_1-1])df} \le \frac{y_1(1-\hat{p}_{mle})(y_1-1)}{2\hat{p}_{mle}}]}{P[F_{(4,2y_1)df} \le \frac{y_1^{\,2}(1-\hat{p}_{mle})}{\hat{p}_{mle}}]}, \quad (67)$$

where $\hat{p}_{mle} = \bar{y}_2$. The *Tail Value at Risk* (TVaR) is

$$TVaR_{Y_3} = E[Y_3 | Y_3 \ge 1, p, y_1] \approx 1 + \frac{y_1 p}{(1-p)^2 P[F_{(2r,2y_1)df} \le \frac{y_1^{\,2}(1-p)}{p}]}. \quad (68)$$