Motivation

Millions of people suffer due to volcanic gases worldwide. Health hazards in volcanic gases [1] like SO₂, H₂S and CO₂ cause fatalities from asphyxiation [2]. Chronic exposure to H₂S increases respiratory diseases [3, 4]. Natural hazards like the 14th April 2010 eruption of Eyjafjallajökull volcano, Iceland cause global public health hazards within a radius of distance from its epi-center in several directions of measurable angle due to wind [5, 6]. The metallic and heavy substances in the ash are trigger illness [4]. Learning from such natural calamity data may not help to prevent it but surely assists to reduce health damages [7].

Developing an appropriate model for the data is a starting point. Modeling such trivariate volcanic data has been a challenge to those who wish to analyze and interpret data evidence [8]. A reason is that the variables are seemingly independent but are correlated otherwise. According to the data (Table 1, Figures 1 and 2). This is a conflict. Such a conflict is not unique to volcanic data analysis but also in tsunami, cyclone, earthquake, and cancer data analysis. In global healthcare data analysis such as the consequences due to volcanic eruptions, the three main random variables (ash in proportion, spilled distance in kilometer, and angle of wind direction) are seemingly independent but therefore data indicate correlations otherwise. This apparent conflict requires a careful scrutiny and selection of appropriate underlying model for the data in more realistic sense rather than the analyst’s convenience. For this purpose, a trivariate probability density function (PDF) with a bonding function of the variables and a flexing parameter, is introduced and investigated with its statistical properties in this article. The flexing parameter allows the PDF to explain the so called “sampling bias” in the collected data. The locations which received more debris have higher chance of being included in the data collection and it is what the statistics literature calls “sampling bias.” The flexing parameter enriches the versatility of the model. Such a novelty based new PDF is named “flexing and bonding trivariate distribution (FBTD)” in this article. Several expressions for FBTD are derived and illustrated using data about the eruption of Eyjafjallajökull volcano, Iceland on 14th April 2010 as it disrupted flights, voyages, travels, and life itself not only in Europe but also elsewhere.

Keywords: Probability; Model; Parameters; Correlation; Data analysis; Maximum likelihood

Summary

In global healthcare data analysis such as the consequences due to volcanic eruptions, the three main random variables (ash in proportion, spilled distance in kilometer, and angle of wind direction) are seemingly independent but their data indicate correlations otherwise. This apparent conflict requires a careful scrutiny and selection of appropriate underlying model for the data in more realistic sense rather than the analyst’s convenience. For this purpose, a trivariate probability density function (PDF) with a bonding function of the variables and a flexing parameter, is introduced and investigated with its statistical properties in this article. The flexing parameter allows the PDF to explain the so called sampling bias in the collected data. The locations which received more debris have higher chance of being included in the data collection and it is what the statistics literature calls “sampling bias.” The flexing parameter enriches the versatility of the model. Such a novelty based new PDF is named “flexing and bonding trivariate distribution (FBTD)” in this article. Several expressions for FBTD are derived and illustrated using data about the eruption of Eyjafjallajökull volcano, Iceland on 14th April 2010 as it disrupted flights, voyages, travels, and life itself not only in Europe but also elsewhere.

Received: May 16, 2017 | Published: July 25, 2017
Table 1: Volcanic eruption of Eyjafjallajökull during 14th April - 13th May 2010 [5].

| Day       | θ (Wind Direction In Angle) | d (Distance (In Kilo Meter) Where Ashes are Found) | y (Percent Ashes More Than 31 µm Less Than 63 µm) |
|-----------|----------------------------|--------------------------------------------------|-----------------------------------------------|
| 14 Apr 2010 | 90                         | 1                                                | 21                                            |
| 14 Apr 2010 | 90                         | 2                                                | 24                                            |
| 14 Apr 2010 | 90                         | 10                                               | 13                                            |
| 14 Apr 2010 | 90                         | 10                                               | 17                                            |
| 15 Apr 2010 | 90                         | 58                                               | 44                                            |
| 15 Apr 2010 | 90                         | 60                                               | 56                                            |
| 15 Apr 2010 | 90                         | 58                                               | 70                                            |
| 15 Apr 2010 | 90                         | 56                                               | 65                                            |
| 16 Apr 2010 | 90                         | 21                                               | 26                                            |
| 16 Apr 2010 | 90                         | 11                                               | 47                                            |
| 22 Apr 2010 | 135                        | 4                                                | 7                                             |
| 5 May 2010  | 135                        | 30                                               | 46                                            |
| 8 May 2010  | 135                        | 13                                               | 12                                            |
| 10 May 2010 | 135                        | 13                                               | 12                                            |
| 13 May 2010 | 135                        | 10                                               | 38                                            |
| 13 May 2010 | 225                        | 14                                               | 10                                            |
| 14 May 2010 | 135                        | 8                                                | 42                                            |
| Average    | 1295.4                     | 22.3                                             | 32.3                                          |
| Variance   | 462.2                      | 407.7                                            |                                               |

Flexing Parameter: $\phi_{\text{shl}} = 1.57$

Figure 1: Box plots of distance and percent ashes at a given wind direction.

Figure 2: The 3-dimensional inter-relations of $y, d, \theta$.

The variance of the distance is $Var(d|\alpha, \beta) = \beta E(d|\alpha, \beta)$, where the expected distance is $E(d|\alpha, \beta) = \alpha \beta$. The parameter $\beta$ captures the proportionality of the expected amount in variance. Furthermore, the entropy $-E \ln f(d|\alpha, \beta)$ of the distance is minimally “$\ln \beta$” but it increases at a rate $\alpha + \ln \Gamma(\alpha) + (1-\alpha)\psi(\alpha)$, where $\psi(\alpha)$ is the well-known digamma function. The parameter $\alpha$ portrays the increment.

The variance of the angle is $Var(\theta) = \frac{[E(\theta)]^2}{3}$, where the
expected angle of the wind direction is $E(\theta)=\pi$. The entropy
$-E \ln f(\theta)$ of the angle is $E(2\pi)$.

The variance of the percent ash is

$$\text{Var}(y)=E(y^2)+\left[1-E(y)\right]^2,$$

where the expected is $E(y)=\frac{\delta}{\delta+y}$.

The entropy $-E \ln f(y)$ of the percent ash spread is

$$E(y)\psi(\delta+y)\psi(\delta-y)+(\delta+y)\psi(\delta+y)-\psi(\delta+y)\psi(\delta-y).$$

In other words, $f(y,\theta,\phi)=p(y)q(\theta)r(d)$, where $p(y), q(\theta)$

Table 2: Correlation among the three random variables.

| Variable                  | Wind Direction (In Angle) | Distance (In Kilo Meter) Where Ashes are Found | Percent Ashes (Mass More Than 31) |
|---------------------------|---------------------------|-----------------------------------------------|----------------------------------|
| Wind direction (in angle) | 1                         | -0.35 (p value = 0.12)                        | -0.44 (p value = 0.05)           |
| Distance (in Kilo Meter)  |                           | 1                                              | 0.77 (p value = 0.0001)          |
| Percent ashes (mass more  | -0.44                      |                                               | 1                                |

This trivariate PDF (1) is new to the literature and hence, it is
named **flexing and bonding trivariate distribution (FBTD)**. The
statistical properties of FBTD are done first in Section 2 and are
illustrated later in Section 3. Final comments are made in Section 4
[10].

A New Trivariate Probability Density and Its Properties

To be realistic, the data collection process is sometimes tilted
unevenly in the collection of natural calamities such as volcanic
eruptions. The tilted sampling process is recognized as
**length-biased sampling** with a weight factor $w(d,\theta)$ in literature.
Then, what is an appropriate weight factor in our scenario? A
rationality for selecting the weight factor is the following. The
area in which the volcanic debris is found is proportional to
the circular circumference $2\pi d$ with radius distance $d>0$. Such
proportionality is well connected to an angle, $\theta>0$ due to wind
direction and hence, it is $\frac{\theta}{2\pi} d=\theta d$.

In addition to this proportionality in the weight factor,
a flexibility to condense or expand the proportionality is
needed and it is done by introducing a finite and non-negative
**flexible parameter** $\phi\geq 0$ so the weight function becomes

$$w(\phi, d, \theta)= (1+\phi \theta d)$$

accompanied the PDF $f(d, \theta)$. Because of the third
variable $Y$, the sampling bias weight function is expanded to
$w(\phi, Y, d)= (1+\phi Y d)$. In other words, the trivariate PDF of the
percent ashes, radius distance, and angle of wind in the collected
data is

and $\phi(d)$ denote respectively marginal PDF of the data variables
$Y$, $d$, and $\theta$. Shanmugam and Chattamvelli [9] for derivations and
statistical details about beta, gamma and uniform distributions.

On the contrary to a seeming impression that the three random
variables $Y, d$ and $\theta$ are independent, their data (Table 2) exhibit
correlated, simply negating the assumption of their independence.
Such a data based due warrants a necessity to derive a realistic
trivariate PDF for the collected data. This necessity results in an
innovative and realistic model with a **bonding function** $w(\phi, Y, d, \theta)$
in which $\phi>0$ is recognized as a **flexing parameter** for the sake of
versatility as it is done in this article.

Citation: Shanmugam R, Singh KP (2017) Flexing and Bonding With A Trivariate Probability Density to Explain Health Consequences of Hazardous
Volcanic Eruptions. Biom Biostat Int J 6(2): 00164. DOI: 10.15406/bbij.2017.06.00164

Copyright: ©2017 Shanmugam et al.
Note that, with \( m = n = p \), the expression (2) is one as it should be. With \( m = 1 = n = p \), the trivariate product moment, \( E(yd\theta) \) is obtained and it is

\[
E(yd\theta) = E(y)E(d)E(\theta) - \frac{\phi \theta}{3(1+\gamma+\theta)} - \frac{4\phi E(\theta)[1+\beta+y(d)]}{3(1+\gamma+\theta)} - \frac{4\phi E(\theta)\beta E(d)}{3(1+\gamma+\theta)}.
\]

In the absence of flexibility or equivalently referring independence among the three random variables (that is, \( \phi = 0 \)), the product moment (3) breaks up to a product of their marginal moments. The expected amount, \( \bar{E}(yd\theta) \) in (3) is at its base \( E(y)dE(d)E(\theta) \) when \( \phi = 0 \) and later increases at a rate

\[
\bar{E}(yd\theta) = \frac{4\phi E(\theta)[1+\beta+y(d)]}{3(1+\gamma+\theta)} - \frac{4\phi E(\theta)\beta E(d)}{3(1+\gamma+\theta)} - \frac{\phi \theta}{3(1+\gamma+\theta)},
\]

when \( \phi \neq 0 \). We define, in this article, the trivariate product variance \( Var(yd\theta) \) as \( Var(yd\theta) = E(y^2d^2\theta^2) - \bar{E}(yd\theta)^2 \). Using (2) and (3), we obtain that

\[
Var(yd\theta) = \frac{4\pi^2\beta^2\alpha\delta}{(\gamma+1)\alpha\beta\delta} \left( \frac{(1+\gamma)(1+\delta)}{(1+\gamma+\theta)} + \frac{\phi \theta}{3(1+\gamma+\theta)} \right) - \frac{2\phi \theta}{3(1+\gamma+\theta)} - \frac{2\phi \delta}{3(1+\gamma+\theta)},
\]

where \( \phi = 0 \) and it later changes when \( \phi \neq 0 \). The predictability becomes less precise when the variance is more and vice versa.

Of interest to healthcare researchers is of course the ability to predict one among the three data variables: \( y, d, \theta \) based on patterns in the other two variables. This requires configuring their conditional PDFs. Suppose a healthcare researcher at a known distance \( D = d \) from the epicenter of a volcano with an observable wind direction \( \theta = \theta \) wonders about receiving an average amount of ash. For this purpose, the conditional PDF, \( f(y|\theta, d) \) is needed. That is,

\[
f(y|\theta, d) = \frac{f(y, \theta, d)}{f(\theta, d)} = \frac{\Gamma(\gamma+\delta+1)(1+\phi y\theta d)}{\Gamma(\gamma+1)(1+\phi y\theta d)\beta^\gamma(1-y)^{\gamma-1}(1-y)^{\gamma-1}},
\]

where \( \phi = 0 \) and it later changes when \( \phi \neq 0 \). The conditional average predictive percent, \( E(y|\theta, d) \) of the ashes based on known wind direction angle, \( \theta \) and the distance, \( d \) is more than the unconditional average predictive percent, \( E(y) \) of ashes without knowing wind direction and location distance. Likewise, we notice that \( Var(y|\theta, d) \leq Var(y) \). The implication is that the conditional average predictive percent of the ashes based on known wind direction angle, \( \theta \) and the distance, \( d \) is more precise (because lesser variance means more precision) than the unconditional average predictive percent of the ashes without knowing the wind direction and location distance.

Agencies responsible to protect the public healthcare often want to project the expected distance, \( E(d|\theta, y) \) based on knowing the angle, \( \theta \) of the wind direction and the percent \( y \) of spreading ashes. This requires configuring the conditional PDF, \( f(d|\theta, y) \) of the distance \( D = d \) from the epicenter of a volcano with an observable wind direction \( \theta = \theta \) and measurable percent of the volcanic ash, \( Y = y \) and it is,

\[
f(d|\theta, y) = \frac{f(y, \theta, d)}{f(\theta, y)} = \frac{(1+\phi y\theta d)}{(1+\phi \beta y\theta d)} \frac{d}{\beta^\gamma e^\gamma e^{d-\gamma-1}}.
\]

which is greater than one. It means that \( E(d|\theta, y) > E(d) \). What does it imply? The conditional average predicted distance, \( E(d|\theta, y) \) for the ashes based on the known wind direction angle, \( \theta \) and the perceived percent, \( y \) of ashes is more than the unconditional average predictive distance, \( E(d) \) of the ashes without knowing wind direction and the percent of ashes spreading. Likewise, we notice that

\[
Var(d|\theta, y) = Var(d) \frac{1+\phi y\beta \theta d(1+\phi y\beta \theta d)}{(1+\phi \beta y\theta d)^2}.
\]

Imposing \( Var(y|\theta, d) \geq Var(y) \). The conditional average projected distance to receive ashes based on known wind direction angle, \( \theta \) and the percent of spreading ashes, \( y \) is less precise (because more variance means lesser precision) than the unconditional average projected distance to receive ashes without knowing wind direction angle and the percent of spreading ashes, \( y \).

Proceeding likewise, having already observed a percent, \( y \) of the volcanic ashes at a known distance, \( d \) from the epicenter of the volcano, an environmental researcher could have done an educated guess of the angle of wind direction on the eruption day. For this purpose, the conditional PDF, \( f(\theta|y, d) \) of the angle is needed and it is

\[
f(\theta|y, d) = \frac{f(y, \theta, d)}{f(y, d)} = \frac{(1+\phi y\theta d)}{2\pi(1+\phi y\theta d)},
\]

where \( 0 < \theta < 2\pi; y, d, \pi > 0 \).
The educated guess, \( E(\theta|y,d) \) of the angle of the wind direction starts at baseline \( E(y)=\pi \) with an adjustment \( \frac{E(\theta|y,d)}{E(\theta)} \) which is greater than one, meaning that \( E(\theta|y,d)>E(\theta) \). What does it imply? The educated conditional average guess, \( E(\theta|y,d) \) of the angle based on known percent ashes, \( y \) at location distance, \( d \) is more than the unconditional average guess angle, \( E(\theta) \) of wind direction without knowing percent of ashes at a location distance, \( d \). Furthermore, we notice that

\[
\text{Var}(\theta|y,d)=\text{Var}(\theta) \left(1+\frac{2\phi y d \pi}{3}\right)^{-1}
\]

implying \( \text{Var}(\theta|y,d)\leq\text{Var}(\theta) \). The educated average guess of the angle for wind direction based on known percent, \( y \) of ashes at distance, \( d \) is more precise than the uneducated average guess of wind direction without knowing percent ashes at location of distance \( d \).

We now proceed to estimate the model parameters from a collected data. Consider a random sample \( y_i,\theta_i,d_i \) of size \( n \geq 2 \) from FDTD (1). Let \( f(y_i,\theta_i,d_i,\phi) \) and \( s_{d_i}^2,s_{\phi}^2 \) denote respectively their sample average and variance. The log-likelihood function is

\[
L(n,y,\theta,d,\phi)=\sum_{i=1}^{n} \ln f(y_i,\theta_i,d_i,\phi).
\]

Then, their maximum likelihood estimators (MLE) are the simultaneous solutions of the score functions \( \hat{\partial}_L(n,y,\theta,d,\phi)=0, \hat{\partial}_u L(n,y,\theta,d,\phi)=0, \hat{\partial}_d L(n,y,\theta,d,\phi)=0 \). They yield

\[
\hat{\phi}_{\text{mle}}=\frac{\sum_{i=1}^{n} y_i \phi y d \pi}{\sum_{i=1}^{n} (y_i \phi y d \pi)^2} - n(\hat{\tau},\hat{\alpha},\hat{\beta},\hat{\delta})^2.
\]

(5a)

where

\[
S = \frac{y_i \phi y d \pi}{(y_i \phi y d \pi)^2} - (\hat{\tau},\hat{\alpha},\hat{\beta},\hat{\delta})^2.
\]

is the standard error of the product variables.
Illustration With Volcanic Data

The 39-day long eruption of Eyjafjallajökull volcano, Iceland started on 14th April 2010 and continued on to 13th May 2010. The ash was seen as far as 100 Kilo Meters in several directions from 90° to 225° due to the wind. The percent of ashes varied up to 63 μm (Table 1). Several jet-engine powered aircrafts were strained, not able to fly as they were exposed to volcanic ash. Due to the prevailing winds, the ash spread out mostly to the southeast, and south, closing the entire European air space for several days. The health hazards were rising among the residents in Europe and even beyond. Exposure to the volcanic ash had adverse effects on respiratory health. The short-term effects of the ash exposure were irritation in eyes, upper airway, or asthma type discomforts. Table 1 for observations on the wind’s angle, the distance up to which the debris is spread out, and the percent of ashes on the sampled days. Their averages and variances are calculated and displayed in Table 1. They warn the existence of their unpredictability. The Table 2 points out their non-zero correlations, emphasizing the importance of bonding and flexing the data variables. Especially, the correlation (0.77) between the distance in kilo meter and the percent of ashes is significant at a p-value 0.0001, emphasizing the importance bonding and flexing the data variables.

The initial parametric values, according to the displayed data averages and variances in Table 1, are obtained. They are \( \hat{\theta}_o = 225° \), \( \hat{\sigma}_o = 11.38 \), \( \hat{\beta}_o = 1.96 \), \( \hat{\gamma}_o = 4.36 \), \( \hat{\phi}_o = 2.95 \). From (5.a), we obtain that \( \phi_o = 1.57 \) which is much higher than zero. It is no wonder the three data variables are not independent. The MLE of the model parameters, from (5.b) through (5.f) are \( \hat{\theta}_{mle} = 194.7° \), \( \hat{\sigma}_{mle} = 9.54 \), \( \hat{\beta}_{mle} = 1.02 \), \( \hat{\gamma}_{mle} = 4.14 \), \( \hat{\phi}_{mle} = 2.01 \). Using the expression (6), an approximate 95% confidence interval is computed and it is \( \phi = (1.57 \pm 0.000002) \). A measure of goodness-of-fit may be based on the log-likelihood statistic. Then one may test the hypothesis \( H_0: \phi = 0 \) against \( H_1: \phi \neq 0 \) by using the likelihood ratio test.

Comments and Conclusions

In some instances but not always, the percent of ashes might proportionally decrease as the distance increases only if the wind is stable. Otherwise, the percent of ashes might be more in a faraway distance if the wind is blowing at an angle and strongly. A realistic probability model is identified and applied to volcanic data analysis. In the model, a parametric symptom is introduced and utilized in this article to check out whether or not the wind direction, distance for volcanic debris, and the percent of ashes have functioned as disconnected entities in a particular natural disaster. With a new and novel flexing and bonding trivariate probability density function (1), we have proved that the three data variables (namely, ash in proportion, spilled distance in kilo meter, and angle direction of the wind) are not independent because \( \phi_{mle} = 1.57 \) which is much higher than zero. Recall that \( \phi = 0 \) is a requirement for the three data variables to be independent.

Acknowledgement

The authors appreciate and thanks an anonymous referee for valuable suggestions which helped to improve contents and quality of the article.

Conflict of Interest

None

References

1. Beaubien SE, Ciotoli G, Lucchese R (2003) Carbon dioxide and radon gas hazard in the Alban Hills area (central Italy). J Vokanol Geotherm Res 123: 63-80. 2. Simkin T, Siebert L, Blong R (2001) Volcano Fatalities-Lessons from the historical record. Science 291(5502): 255. 3. Annuziatellis A, Ciotoli, GLombardi S (2003) Short- and long-term gas hazard: the release of toxic gases in the Alban Hills volcanic area (central Italy). J Geochem Exploration 77: 93-108. 4. Small D, Naumann T (2001) The global distribution of human population and recent volcanism. Environ Haz 3: 93-109. 5. Guslason SR, Alfredsson HA, Eiríksdóttir ES, Hasenckam T, Stipp SLS (2011) Volcanic ash from the 2010 Eyjafjallajökull eruption. Applied Geochemistry 26: S188-S190. 6. Grattan J, Durand M, Gilbertson D (2003) Human sickness and mortality rates in relation to the distant eruption of volcanic gases: rural England and the 1783 eruption of the Laki fissure, Iceland. In: Geology and Health: Closing the Gap, Skinner HCW & Berger A (Ed.) Oxford: Oxford University Press, USA. p. 19-24. 7. World Health Organization (2010) Volcanic Ash Cloud over Europe. 8. Durand M, Grattan J (2001) Effects of volcanic air pollution on health. Lancet 357: 164. 9. Shanmugam R, Chattamvelli R (2015) Statistics for Scientists and Engineers, John Wiley Press, 111 River Street, Hoboken, NJ 07030-5774, ISBN: 978-1-118-228968. 10. Baxter PJ (1990) Medical effects of volcanic eruptions. Bull Vokanol S2: 532-544.