Accident Analysis and Prevention 107 (2017) 164–172

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Full length article

Analysis of multiple tank car releases in train accidents

Xiang Liu[a,*], Chang Liu[b], Yili Hong[c]

[a] Department of Civil and Environmental Engineering, Rutgers University, United States
[b] Department of Statistics, Rutgers University, United States
[c] Department of Statistics, Virginia Tech, United States

ARTICLE INFO

Keywords:
Hazardous materials
Safety
Tank car release
Correlated binomial model
Transportation risk

ABSTRACT

There are annually over two million carloads of hazardous materials transported by rail in the United States. The American railroads use large blocks of tank cars to transport petroleum crude oil and other flammable liquids from production to consumption sites. Being different from roadway transport of hazardous materials, a train accident can potentially result in the derailment and release of multiple tank cars, which may result in significant consequences. The prior literature predominantly assumes that the occurrence of multiple tank car releases in a train accident is a series of independent Bernoulli processes, and thus uses the binomial distribution to estimate the total number of tank car releases given the number of tank cars derailing or damaged. This paper shows that the traditional binomial model can incorrectly estimate multiple tank car release probability by magnitudes in certain circumstances, thereby significantly affecting railroad safety and risk analysis. To bridge this knowledge gap, this paper proposes a novel, alternative Correlated Binomial (CB) model that accounts for the possible correlations of multiple tank car releases in the same train. We test three distinct correlation structures in the CB model, and find that they all outperform the conventional binomial model based on empirical tank car accident data. The analysis shows that considering tank car release correlations would result in a significantly improved fit of the empirical data than otherwise. Consequently, it is prudent to consider alternative modeling techniques when analyzing the probability of multiple tank car releases in railroad accidents.

1. Introduction

Each year, over two million carloads of hazardous materials (hazmat) are transported by American railroads (AAR, 2017). Although hazardous materials accounts for only 7% of U.S. rail traffic, it is responsible for a major share of railroads' liability and insurance risk (AAR, 2017). Since 2005, the shale oil production boom in North America has led to significant growth in rail transport of flammable liquids. Being different from roadway transport of hazardous materials, a train can carry multiple tank cars, sometimes over 100 tank cars in a single train. Therefore, a train accident has the potential to cause the derailments and releases of multiple tank cars. Several recent multiple-tank-car release incidents, particularly the derailments in Lac-Mégantic, Canada in July 2013, Aliceville, Alabama in November 2013, and Casselton, North Dakota in December 2013, all underscore the vital importance of understanding and preventing multiple-car release risk (Liu et al., 2014; Liu, 2017).

One principal task in railroad hazmat transportation risk management is to understand the number of tank cars releasing per train accident. Previous studies predominantly assumed that tank car releases per train accident are mutually independent. Under this assumption, binomial distribution has been used to estimate the number of tank cars releasing given the total number of tank cars derailed (e.g. Nayak et al., 1983; Glickman et al., 2007; Bagheri et al., 2011, 2012, 2014; Liu et al., 2014). To our knowledge, Liu and Hong (2015) was the only published study that accounts for the dependency between tank car releases in the same accident. They found that Beta Binomial model outperforms the traditional binomial model based on one empirical dataset. Their study finds that accounting for tank car release dependency could substantially change risk estimation for the incidents involving a large number of tank cars releasing contents. Therefore, an accurate estimation of multiple tank car release probability is very critical for railroad hazardous materials risk management.

However, Liu and Hong (2015) paper has two major limitations. First, only one type of dependency structure is considered. It is worth investigating whether other dependency structures could further improve the fit of the empirical data. Second, they focused on modeling the conditional mean value of the number of tank cars releasing per accident. In addition to the conditional mean, other distributional statistics (e.g. median, 80th percentile) are also worth investigation,

* Corresponding author.
E-mail address: xiang.liu@rutgers.edu (X. Liu).

http://dx.doi.org/10.1016/j.aap.2017.07.007
Received 16 May 2017; Received in revised form 5 July 2017; Accepted 5 July 2017

especially when the distribution of tank cars releasing is asymmetrical.

This paper aims to advance railroad hazardous materials transportation risk analysis, with the following two objectives:

- Modeling multiple tank car release probability using alternative correlated binomial models (including Beta Binomial model, Increment Risk model and Family History model, respectively), in comparison with the binomial model that assumes no dependency between tank car releases
- Explore the use of quantile statistics to measure the severity of a railroad hazmat release incident, in addition to using the conditional mean

This research focuses on the releases caused by mechanical damage incurred by tank cars in train accidents, without accounting for the releases resulting from thermal tearing, which is a process by which a fire impinging on the tank causes the steel to weaken. Accounting for thermal-tearing-caused tank car release risk is the next step of this work.

The paper is structured as follows. Section 2 presents a review of the literature and clarifies the intended contributions of this paper. Section 3 introduces the statistical methodology that is comprised of three types of correlated binomial models. Section 4 presents the data used for statistical modeling. Sections 5 to 7 discuss the results and implications to railroad safety analysis. Sections 8 and 9 conclude the study and suggest possible future research directions.

2. Literature review

Tank cars today are the second most common type of railroad freight car in North America, accounting for approximately 20 percent of the rail car fleet (Barkan et al., 2013). Tank cars annually transport over two million shipments of hazardous materials that are essential to the nation's economy (Barkan et al., 2013).

The Railway Supply Institute (RSI) and the Association of American Railroads (AAR) developed industry-wide tank car accident statistics since the 1970s (Treichel et al., 2006). Using this database, the AAR-RSI published statistics regarding the safety performance of a tank car by its safety design. For example, if a non-jacketed 111A100W1 (7/16 inch tank thickness) derails, its release probability is 0.196. By contrast, the release probability of a jacketed CPC-1232 car (7/16 inch tank thickness) is reduced to 0.046. Note that the published AAR-RSI tank car accident statistics focus on single tank cars, without accounting for the possible correlation between multiple tank car releases within the same train accident.

In railroad hazmat transportation risk analysis, the estimation of the number of tank cars releasing is a pivotal task. Differing from roadway transport of hazardous materials, a train accident can potentially cause the derailment and releases of multiple tank cars. Given the total number of tank cars derailed, the number of tank cars releasing hazardous materials follows a probabilistic distribution, depending on whether tank car releases within the same accident are independent:

a. Derailed tank cars have independent release probabilities. Almost all previous studies were based on this assumption and they used a binomial distribution to estimate the total number of tank car releases given the number of tank cars derailed (e.g. Nayak et al., 1983; Glickman et al., 2007; Bagheri et al., 2011, 2012, 2014).
b. Derailed tank cars in the same accident have correlated release probabilities. This scenario accounts for the interactions among tank car release probabilities within the same train accident. To our knowledge, the only published study addressing this scenario was presented by Liu and Hong (2015). They used a Beta Binomial model to describe a specific correlation structure between releasing tank cars, and found that the Beta Binomial model outperformed the traditional binomial model.

While Liu and Hong (2015)'s study indicates the promise of fitting the tank car accident data by accounting for the correlations of tank car releases within the same train accident, there are still a number of unexplored questions, including at least the following:

- Would different correlation structures have different fits of the empirical tank car safety data?
- Does a particular model always have a better performance than another model, or is the model performance is dependent on the specific dataset?
- How do we measure the severity of a railroad tank car release incident? Do we use the conditional mean value or quantile statistics? How would these statistics vary in different statistical models?

This paper is intended to establish a new methodological framework for analyzing tank car releases based on historical railroad tank car accident data. In particular, we consider three alternative correlated binomial models, including Beta Binomial (BB) model, Family History (FH) model and Increment Risk (IR) model, respectively. Two independent sample datasets are used to validate and compare the performance of these models, versus the conventional binomial model. Finally, based on the model output, we analyze the mean value and quantile statistics of the probabilistic distribution of the number of tank cars releasing per train accident.

3. Statistical methodology

Derailment is a common type of freight-train accident in the United States (Liu et al., 2012; Liu, 2016). Therefore, this paper focuses on derailment-caused tank car releases. Let D_i denote the release of the i^{th} derailed tank car in a train derailment ($D_i = 1$ if this car releases and 0 otherwise). Let P_i denote its release probability (also called Bernoulli probability). As a result, the total number of tank cars releasing (denote as Y_n) given n tank cars derailed in a freight-train derailment can be expressed as:

$$Y_n = \sum_{i=1}^{n} D_i \tag{1}$$

The release of a derailed tank car can be viewed as a Bernoulli variable. It can be assumed that the Bernoulli indicators D_i are dependent in such a way that the conditional probability of release in any tank car releasing depends on the total number of cars releasing prior to the particular tank car. As described in Liu and Hong (2015), this assumption seems to be reasonable given the fact that the total number of cars releasing reflects the total accident kinetic energy, which is related to tank car release probability (Liu et al., 2014).

Mathematically, the above-mentioned dependency assumption is expressed as follows:

$$\mathrm{P}(D_i = 1|D_1, D_2, \ldots, D_{i-1}) = P(D_i = 1|D_1 + D_2 + \ldots + D_{i-1}) \tag{2}$$

For illustrative convenience, we adopt a more concise notion of tank car release dependency based on a previous statistical study from Yu and Zelterman (2002):

$$C_n(s) = P(D_n = 1|D_1 + D_2 + \ldots + D_{n-1} = s) \tag{3}$$

where $C_n(s)$ denotes the conditional probability that the nth derailed tank car would release, given that there are s tank cars releasing prior to it. We also define $C_1 = C_1(0) = P(D_1 = 1)$. Let $P_n(s)$ ($n \geq 1$) denotes the probability of releasing s tank cars out of n derailed tank cars, that is

$$P_n(s) = P(D_1 + \ldots + D_n = s) \tag{4}$$

Using the Law of Total Probability (LTP), we can derive P_n using the following recursive algorithm:

$$P_n(s) = C_n(s-1)P_{n-1}(s-1) + [1 - C_n(s)]P_{n-1}(s) \tag{5}$$

Eq. (5) provides a recursive algorithm to calculate the probability mass function (PMF) of the number of dependent tank car releases (s) given the number of tank car derailed (n) per train derailment. Based on Eq. (5), we can also calculate the probability of the two extreme scenarios, which are either all derailed tank cars release $P_n(n)$, or none of the derailed cars release $P_n(0)$.

$$P_n(n) = \prod_{i=0}^{n-1} C_{i+1}(i) \tag{6}$$

$$P_n(0) = \prod_{i=0}^{n-1} (1 - C_{i+1}(0)) \tag{7}$$

If the conditional probability $C_n(s)$ is not dependent on the total number of cars derailed (n), we can simplify $C_n(s) = C(s)$. Accordingly, $P_n(s)$ can be written as

$$P_n(s) = \left\{\prod_{j=0}^{s-1} C(j)\right\}\left\{\sum_X \prod_{j=0}^{S} (1 - C(j))^{x_j}\right\} \text{ where } X = X_n(s) = \{x_j = 0,1,...;j = 0,1,...,s; \sum_j x_j = n - s\} \tag{8}$$

The conditional probability $C_n(s)$ can have different forms, each of which can lead to a different correlated binomial model. This paper considers three correlated binomial models based on different $C_n(s)$ functions. Appendix A will discuss each alternative model.

In the three correlated binomial models, the parameters can be fitted to the empirical data using the method of Maximized Likelihood (ML). Using the Family History model as an example, its parameters p and p' can be derived by maximizing the likelihood function:

$$(p, p') = \arg_{p,p'} \max \prod_k \ln[P(Y_k)] \tag{9}$$

where (p, p') are parameters in the Family History model and $P(Y_k)$ is the probability of releasing Y_k tank cars in the kth train accident. Similarly, the parameters in other models can be derived accordingly (note: the Family History model will be introduced in Appendix A).

4. Model development and comparison using empirical data

To illustrate the application of our methodology, we assembled two independent, random samples of hazardous material train derailments between 1990 and 2010 based on data from the US railroad industry. The datasets used for the statistical analysis are presented in Appendix B. The purpose of using multiple random samples is to understand whether the "best" statistical model depends on the selected data. In those train derailments, all the derailed tank cars conform to non-jacketed DOT 111A100W1 tank car design features (7/16 inch tank thickness and head thickness). This was one of the common hazardous materials tank cars used in North America (Liu and Hong 2015). Derailment speed is approximately 30 mph. Each train accident resulted in 10 railcars derailed (including both hazmat cars and non-hazmat cars). The selection of "homogeneous" accident conditions could better isolate the effect of tank car release dependency by controlling other factors constant. The likelihood values of the four models with parameters estimation from two data sets are shown below.

A higher likelihood value indicates a better fit of the empirical data. For the first dataset, the likelihood of Beta Binomial model is the highest among the four models (Table 1). However, for the second dataset, the Increment Risk model has the highest likelihood. In both datasets, the binomial model has the lowest likelihood. It indicates that the empirical data exhibits interdependency among tank car releases.

Besides comparing alternative correlated binomial models, we also used a likelihood ratio (LR) test to examine whether the correlation parameter is significant in the Beta Binomial model, Family History model and Increment Risk model, each compared to the benchmark binomial model. The detailed process is as follow. For illustrative convenience, we only use Family History model as an example:

Null Hypothesis $H_0 : p = p'$ (no tank car release correlation within the same train accident).

Alternative Hypothesis $H_1 : p \neq p'$ (there is tank car release correlation within the same train accident)

To test the hypothesis, a statistic called Deviance is calculated as follows:

$$D = 2 \times [\ln L(Family\ History) - \ln L(Binomial)]$$

where D = deviance, ln$L$(*Family History*) = logarithmic likelihood of a Family History model and ln$L$(*Binomial*) = logarithmic likelihood of a binomial model.

According to the statistical theory, the deviance approximately follows a *Chi-square* distribution (Agresti 2007). Using the information from Table 1a and b, we have the deviance and P-value table below in Table 2a and b:

It is found that the correlation parameter in each of the three models is statistically different from zero (P < 0.01), indicating that the necessity of accounting for tank car release interdependencies exhibited from the empirical data. Finally, we conduct a Chi-square goodness-of-fit test to study whether the four models adequately fit the empirical tank car accident data:

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{10}$$

Table 1
(a) Parameter estimates (data set I). (b) Parameter estimates (data set II)

a)

	Binomial	Beta Binomial	Family History	Increment Risk
Parameter Estimation	0.216	(p, α) = (0.222, 0.656)	(p, p') = (0.171, 0.618)	(α, β) = (−1.584, 1.626)
Log-likelihood	−48.60	−43.62	−44.70	−44.64
Likelihood	7.83E − 22	1.13E − 19	3.87E − 20	4.11E − 20

b)

	Binomial	Beta Binomial	Family History	Increment Risk
Parameter Estimation	0.209	(p, α) = (0.178, 0.537)	(p, p') = (0.128, 0.720)	(α, β) = (−1.877, 1.451)
Log-likelihood	−64.39	−52.20	−53.33	−50.56
Likelihood	1.08E − 28	2.14E − 23	6.91E − 24	1.10E − 22

Table 2
(a) Likelihood Ratio Test, Compared to the Binomial Model (Data Set I). (b) Likelihood Ratio Test, Compared to the Binomial Model (Data Set II).

a)

	Family History	Increment Risk	Beta Binomial
Deviance	7.8	7.92	9.96
P-value	0.005	0.005	0.002

b)

	Family History	Increment Risk	Beta Binomial
Deviance	22.12	27.66	24.38
P-value	2.56E-6	2.43E-7	7.91E-7

Table 3
(a) Chi-square goodness-of-fit test (data set I). (b) Pearson Chi-square goodness-of-fit test (data set II).

a)

	Binomial	Beta Binomial	Family History	Increment Risk
Chi-square value	61.2	61.2	65.69	66.52
p-value (DF = 55)	0.26	0.26	0.15	0.14
Conclusion (Type I error = 0.05)	Fit	Fit	Fit	Fit

b)

	Binomial	Beta Binomial	Family History	Increment Risk
Chi-square value	78.86	78.86	68.03	73.48
p-value (DF = 58)	0.03	0.03	0.17	0.08
Conclusion (Type I error = 0.05)	Not fit	Not fit	Fit	Fit

where O_i is the observed number of tank cars releasing per train derailment; and E_i is the expected (mean) number of tank cars releasing per accident using the probabilistic model. For the binomial and Beta Binomial model, their expected means can both be expressed as $E[Y_n] = np$ For the Family History model, the mean of Y_n is $E[Y_n] = np' + (p - p')/p[1 - (1 - p)^n]$. For the Increment Risk model, we don't have a closed form of the mean of Y_n, but we can numerically calculate the expected value of Y_n. Using the raw data and equations above, we conduct Chi-square tests on both empirical data sets. The detailed results are shown in Table 3.

From the above two tables, for data set 1, all of the four models adequately fit the empirical data set. However, for data set 2, neither the binomial or Beta Binomial models would fit the empirical data (they have the same expected values of Y_n). By contrast, the Increment Risk model seems to have the "best" performance based on the likelihood value (Table 1b). All the three correlated binomial models outperform the traditional binomial model based on the likelihood value for both Table 1a and b. Depending on the specific dataset, the proper model can be developed and validated using the methodology presented in this paper.

5. Practical application to railroad safety analysis

The correlated binomial models can be used to predict the probabilistic distribution of the number of tank cars releasing given the total number of tank cars derailed. For example, assuming that there are 10 tank cars derailed, the possible number of tank cars releasing ranges from 0 to 10 (11 scenarios). Fig. 1a and b present the predicted probabilistic distribution of the number of cars releasing out of 10 tank cars derailed, using binomial model, Family History model, Incremental Risk model as well as the Beta Binomial model, respectively. The parameters in Fig. 1a are based on dataset I, and the parameters in Fig. 1b are based on dataset II (Table 3).

From the two figures above, the distribution functions of the four models are quite different. For the Beta Binomial model, the probability of a multiple-car release decreases for an increasing number of tank cars. For the binomial model, it's highly skewed to the small value of tank cars releasing. The other two models, Family History model and Increment Risk model predict a higher probability of a larger number of cars releasing. Both models tend to reflect a large number of tank cars releasing contents.

The significant difference between the binomial model (non-dependency assumption) and the three alternative correlated binomial models has important practical implications. Currently, research has predominantly used the binomial model to estimate the number of tank car releasing in railroad hazmat transportation risk analysis. As shown in both Fig. 1a and b, the binomial model predicts an extremely low probability of a very large number of cars releasing. By contrast, the correlated binomial models, particularly the Incremental Risk model, provides a much higher prediction of a large, multiple-car release incident. For example, in Fig. 1b, the probability that all derailed tank

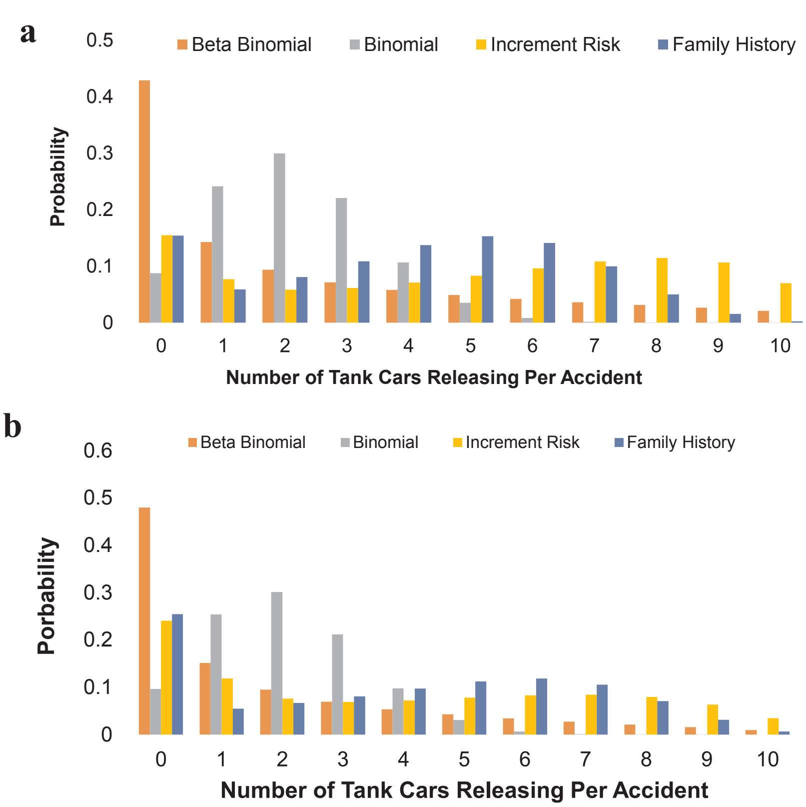

Fig. 1. a Daistribution of tank cars releasing by different models from data set I. (10 tank cars derailed per accident). Notes: The parameter estimation was obtained from Table 1a. (b) Dbistribution of tank cars releasing by different models from data set II.(10 tank cars derailed per accident). Notes: The parameter estimation was obtained from Table 1b.

Table 4
Quantile-based hazmat release severity index. (using dataset I as an example, only 12 records are displayed here for illustration)

ID	N (number of tank cars derailed)	S (number of tank cars releasing)	Binomial	Beta Binomial	Family History	Incremental Risk	Binomial	Beta Binomial	Family History	Incremental Risk
			Median (50% quantile)				80% quantile			
1	1	0	0.0000	0.0000	0.0000	0.0000	0.0750	0.1003	0.0000	0.0000
2	1	1	0.0000	0.0000	0.0000	0.0000	0.0750	0.1003	0.0000	0.0000
3	1	0	0.0000	0.0000	0.0000	0.0000	0.0750	0.1003	0.0000	0.0000
4	1	1	0.0000	0.0000	0.0000	0.0000	0.0750	0.1003	0.0000	0.0000
5	2	0	0.0000	0.0000	0.0000	0.0000	0.5478	0.6068	0.5421	0.4966
6	4	0	0.2945	0.0000	0.1546	0.1248	1.0359	1.4758	1.8326	1.8626
7	1	0	0.0000	0.0000	0.0000	0.0000	0.0750	0.1003	0.0000	0.0000
8	5	0	0.5005	0.0000	0.7328	0.5841	1.4275	1.8951	2.5104	2.8088
9	1	0	0.0000	0.0000	0.0000	0.0000	0.0750	0.1003	0.0000	0.0000
10	3	0	0.0465	0.0000	0.0000	0.0000	0.7993	0.9857	1.1971	1.0127
11	2	1	0.0000	0.0000	0.0000	0.0000	0.5478	0.6068	0.5421	0.4966
12	2	0	0.0000	0.0000	0.0000	0.0000	0.5478	0.6068	0.5421	0.4966

cars release hazardous materials is 1.56E-7 for binomial model (1.56 out of 10 million chance), 0.01 for Beta Binomial model, 0.007 for Family Historical model and 0.035 for Incremental Risk model.

To be prudent in risk analysis, we suggest the analysts to test at least these four models (and other models if proper) to understand 1) which model better fits the empirical data; and 2) the impacts of influencing factors on the corresponding predicted distribution. Besides the four models considered herein, there might be other forms of tank car release probabilistic modeling. When these models and the needed information are available, the methodology procedure in this paper can be adapted accordingly.

6. Quantile index of the number of tank cars releasing per accident

In the probabilistic distribution, certain statistics such as the mean or quantiles can be useful in practice to indicate the “severity” of a hazmat release incident, as measured by the number of tank cars releasing contents. For illustration, we used the empirical dataset I to develop a quantile-based analysis using the median (50% quantile) and 80% quantile, respectively (Table 4). We used the dataset with 56 observations to evaluate the difference of the estimated median values using different correlated binomial methods. Strictly, the probabilistic distribution of the number of tank cars releasing (denote as s) is discrete. In this example, we use a linear function to extrapolate quantile statistics. For instance, suppose the number of total derailed cars is 2 (i.e. n = 2), and $P(s = 0) = 0.4$, $P(s \leq 1) = 0.76$. In this case, there are no integer values for s to attain the probability of 0.5 exactly. As a result, we linearize the cumulative distribution of s between 0 and 1 (0.4 and 0.76), and interpolate a point which has the exact value of 0.5 in the cumulative density function. This linearization approach allows us to obtain any quantile statistics and directly compare the quantile outputs derived from different models.

The quantile statistics provide a complementary way for railroad safety analysts to estimate hazmat train release severity as measured by the number of cars releasing contents. For example, if the 80% quantile statistic is used, we can estimate the maximum number of cars releasing with 80% of the cumulative probability. This would provide additional information to the estimation of hazmat release severity that is solely based on the mean (average) value. For instance, for the 49th incident (ID = 49) in dataset I, 4 tank cars are derailed, among which 2 cars released contents. This empirical severity (2 tank cars releasing contents) is close to the predicted 80th percentile using the Family History model or Incremental Risk model. For both models, the 80th percentile values are around 1.8, indicating that there is 20 percent chance (1-0.8) that the empirical number of tank cars releasing given 4 cars derailed would be higher than 1.8. For the same accident, the mean (average) number of cars releasing is around 1.1 (Appendix A). Again, these parameters were developed based on train accidents with non-jacketed DOT 111A100W1 car, 10 railcars derailed (including non-tank cars), accident speed 30 mph. The model can be adapted to other railroad operating scenarios.

7. Recommendations for railroad hazmat transportation risk analysis

Based on this research, we make the following recommendations for the research community when conducting railroad hazmat transportation risk analysis:

- The two empirical tank car accident datasets exhibit statistically significant correlations among tank car releases within the same train accident. In the presence of multiple-tank-car release correlations, the binomial model in the literature could significantly misestimate the chance of a very large number of cars releasing contents.
- Three alternative correlated binomial models can be adapted to specific datasets. According to the analysis of the two data samples, the Family History model and Increment Risk model tend to predict a much higher probability that all (or most) of the derailed cars release their contents. This is a very important safety concern to both researchers and practitioners in railroad community. The analysis shows that the estimated probability of all-derailed-tank-car-release-contents would vary by magnitudes when using different probabilistic models. This also suggests that analysts should prudently consider alternative statistical modeling approaches to properly analyze tank car safety data. Researchers and practitioners can use the model to assess hazardous materials transportation risk, thereby evaluating and implementing promising risk mitigation strategies, such as tank car safety improvement.
- In addition to the conditional mean (average), researchers can also consider quantile statistics, to measure the severity index of the number of tank car releasing. When the probability distribution of the number of tank car releasing is highly skewed, quantile statistics could provide a more comprehensive view of the output. Particularly, quantile statistics can help analysts better understand tank car release risks in more extreme (or worst case) scenarios (e.g. 80% quantile severity instead of the average severity).

8. Conclusions

This research analyzes the number of tank car releases in a train derailment, using four alternative correlated binomial models. When there exist correlations among tank car releases within the same accident, the three correlated binomial models (i.e., the Beta Binomial model, the Family History model, and the Increment Risk model) have a better fit to the empirical data than the traditional binomial model. Based on the empirical data used in this paper, the Family History model and Increment Risk model may predict a much higher probability of a very large number of tank cars releasing (e.g. all derailed tank cars release contents). Also, we propose the use of quantile statistics (e.g. median, 80% quantile) to evaluate the number of tank cars releasing, in addition to the conditional mean (average). The methodology developed in this paper can be used by railroad analysts to improve the accuracy of risk analysis with respect to rail transport of hazardous materials.

9. Future research

First, there is still lack of knowledge regarding the dissipation of accident kinetic energy and the dynamics of hazardous materials train derailments, as well as the occurrence of thermal-tearing-caused releases in fires. While this paper focuses on the releases caused by mechanical damage incurred by tank cars from a statistical perspective, both physical impact and thermal tearing impact deserve more research in the future. Second, this paper does not account for interactive effects among different types of hazardous materials. Future research can be conducted to model this type of risk when a train carries multiple types of hazardous materials.

Appendix A. —Correlated binomial modeling

(A.1) Family History (FH) model

In the Family History (FH) model, we denote two probability parameters p and p'

$$C_{n+1}(s) = \begin{cases} p & for \quad s = 0, \\ p' & for \quad s = 1, ..., n. \end{cases} \tag{A.1}$$

The model assumes that the conditional probability of a tank car releasing is p, if no other tank cars prior to it release contents. If there is already at least one tank car releasing, then p' is the conditional probability instead. In particular, if p = p', then the Bernoulli variables D_i are independent and their sum Y_n is a binomial distribution. Binomial model is a special case of the FH model. The FH model was originally used in a medical research study by Yu and Zelterman (2002). The FH distribution of Y_n has the form:

$$P(Y_n = s) = P_n(s) = \begin{cases} (1-p)^n & for \quad s = 0, \\ p(p')^{s-1} \sum_{j=0}^{n-s} \binom{n-j-1}{s-1} (1-p)^j (1-p')^{n-s-j} & for \quad s = 1, ..., n. \end{cases} \tag{A.2}$$

Based on Equations (A.1) and (A.2), we can derive the probability distribution of dependent tank cars releasing in the same train derailment. The mean and variance of the number of tank cars releasing are:

$$E[Y_n] = np' + (p - p')/p[1 - (1-p)^n] \tag{A.3}$$

$$\text{var}[Y_n] = np'(1-p') + (p-p')\{(1+2np')p(1-p)^n - (p-p')(1-p)^{2n} - p'\}/p^2 \tag{A.4}$$

The two equations show that the FH model has different mean and variance values, compared to the binomial model.

9.1. Increment Risk (IR) Model

An alternative model can be used when the conditional probability of releasing $C_n(s)$ is a strictly monotone function with respect to the number of tank cars that release contents. This model is called Incremental Risk (IR) model, with the following structure:

$$C_{n+1}(s) = \frac{exp(\alpha + \beta s)}{1 + \exp(\alpha + \beta s)} \tag{A.5}$$

where α and β are two parameters that need to be determined. Binomial model is a special case in which $\beta = 0$. The parameterβ indicates the sign of the dependency between tank car releases. If $\beta > 0$, there is a positive dependency between tank car releases. If $\beta < 0$, it will be a negative dependency. The probability of $P(Y_n = s)$ for this model can be computed numerically from the recursive relation given at Eq. (5). The initial condition of the recursive process can be achieved as below.

$$P_n(0) = \prod_{i=0}^{n-1} \frac{1}{1 + \exp(\alpha)} \tag{A.6}$$

Using Eq. (6), we have

$$p_n(n) = \prod_{i=0}^{n-1} \frac{\exp(\alpha + \beta i)}{1 + \exp(\alpha + \beta i)} \tag{A.7}$$

There is no closed form of the distribution function of the Increment Risk model, we can only calculate it as well as the mean values numerically from Eq. (5).

9.2. Beta Binomial (BB) Model

The third correlated binomial model we used is the Beta Binomial model, which was also the one studied in Liu and Hong (2015). This model considers the total number of tank cars derailed (n) in the tank car release dependency structure. The Beta Binomial model takes the following form:

$$C_{n+1}(s) = \frac{\alpha s + p}{n\alpha + 1} \tag{A.8}$$

where parameter $0 < p < 1$ and $\alpha \geq -p/N$, N is the largest number of derailed tank cars. In this model, the sign of α determines the sign of the dependency between the releasing cars. Obviously, when $\alpha = 0$, it is simplified as the binomial model. Using the recursive relation in (5), we can get the distribution of Y_n is

$$P_n(s) = \binom{n}{s} \prod_{i=0}^{s-1} (i\alpha + p) \prod_{j=0}^{n-s-1} (j\alpha + 1 - p) / \prod_{k=0}^{n-1} (k\alpha + 1) = \binom{n}{s} \frac{\Gamma(s + p/\alpha)\Gamma\{n - s + (1-p)/\alpha\}\Gamma(\alpha^{-1})}{\Gamma(p/\alpha)\Gamma\{(1-p)/\alpha\}\Gamma(n + \alpha^{-1})} \tag{A.9}$$

which is the beta binomial distribution with parameters p/α and $(1-p)/\alpha$. We can also calculate the mean and variance function of the beta binomial model as

$$E[Y_n] = np$$
$$Var[Y_n] = np(1-p)\{1 + (n-1)\alpha/(\alpha+1)\} \tag{A.10}$$

Appendix B. —Tank car release data sets and results

For example, in the 8th accident (Accident ID = 8), there were 5 tank cars derailed (n = 5), and none of them released contents (s = 0). Using the Family History (FH) model, the predicted mean (average) number of tank cars releasing in this accident is 1.50. Also, the estimated probability of having this observation (i.e. none of the five derailed tank cars release contents) is 0.39. These parameters were developed based on train accidents with non-jacketed DOT 111A100W1 car, 10 railcars derailed (including non-tank cars), accident speed 30 mph.

Data Set I (parameter estimators of each model are in Table 1a).

			Predicted Conditional Mean (Average)				Estimated Probability of the Observation			
Accident ID	Number of Tank Cars Derailed	Number of Tank Cars Releasing	Bin	BB	FH	IR	Bin	BB	FH	IR
1	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
2	1	1	0.22	0.22	0.17	0.17	0.22	0.22	0.17	0.17
3	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
4	1	1	0.22	0.22	0.17	0.17	0.22	0.22	0.17	0.17
5	2	0	0.43	0.43	0.42	0.40	0.61	0.67	0.69	0.69
6	4	0	0.86	0.86	1.09	1.10	0.38	0.56	0.47	0.47
7	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
8	5	0	1.08	1.08	1.50	1.57	0.30	0.53	0.39	0.39
9	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
10	3	0	0.65	0.65	0.73	0.70	0.48	0.61	0.57	0.57
11	2	1	0.43	0.43	0.42	0.40	0.34	0.21	0.21	0.22
12	2	0	0.43	0.43	0.42	0.40	0.61	0.67	0.69	0.69
13	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
14	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
15	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
16	6	4	1.30	1.30	1.94	2.13	0.02	0.06	0.14	0.12
17	2	0	0.43	0.43	0.42	0.40	0.61	0.67	0.69	0.69
18	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
19	2	0	0.43	0.43	0.42	0.40	0.61	0.67	0.69	0.69
20	2	0	0.43	0.43	0.42	0.40	0.61	0.67	0.69	0.69
21	1	1	0.22	0.22	0.17	0.17	0.22	0.22	0.17	0.17
22	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
23	3	0	0.65	0.65	0.73	0.70	0.48	0.61	0.57	0.57
24	3	0	0.65	0.65	0.73	0.70	0.48	0.61	0.57	0.57
25	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
26	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
27	1	1	0.22	0.22	0.17	0.17	0.22	0.22	0.17	0.17
28	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
29	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
30	3	0	0.65	0.65	0.73	0.70	0.48	0.61	0.57	0.57
31	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
32	1	1	0.22	0.22	0.17	0.17	0.22	0.22	0.17	0.17
33	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
34	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
35	2	1	0.43	0.43	0.42	0.40	0.34	0.21	0.21	0.22
36	2	1	0.43	0.43	0.42	0.40	0.34	0.21	0.21	0.22
37	2	1	0.43	0.43	0.42	0.40	0.34	0.21	0.21	0.22

38	4	1	0.86	0.86	1.09	1.10	0.42	0.18	0.17	0.21
39	2	0	0.43	0.43	0.42	0.40	0.61	0.67	0.69	0.69
40	3	3	0.65	0.65	0.73	0.70	0.01	0.08	0.07	0.07
41	2	1	0.43	0.43	0.42	0.40	0.34	0.21	0.21	0.22
42	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
43	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
44	1	1	0.22	0.22	0.17	0.17	0.22	0.22	0.17	0.17
45	2	2	0.43	0.43	0.42	0.40	0.05	0.12	0.11	0.09
46	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
47	6	0	1.30	1.30	1.94	2.13	0.23	0.50	0.33	0.33
48	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
49	4	2	0.86	0.86	1.09	1.10	0.17	0.11	0.19	0.14
50	3	0	0.65	0.65	0.73	0.70	0.48	0.61	0.57	0.57
51	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
52	3	0	0.65	0.65	0.73	0.70	0.48	0.61	0.57	0.57
53	1	0	0.22	0.22	0.17	0.17	0.78	0.78	0.83	0.83
54	1	1	0.22	0.22	0.17	0.17	0.22	0.22	0.17	0.17
55	4	0	0.86	0.86	1.09	1.10	0.38	0.56	0.47	0.47
56	4	0	0.86	0.86	1.09	1.10	0.38	0.56	0.47	0.47
Total	111	24								

Notes:
BB = Beta Binomial model.
Bin = Binomial model.
IR = Incremental Risk model.
FH = Family History model.
Data Set II (parameter estimators of each model are in Table 1b).

			Predicted Conditional Mean				Estimated Probability of the Observation			
Accident ID	Number of Tank Cars Derailed	Number of Tank Cars Releasing	Bin	BB	FH	IR	Bin	BB	FH	IR
1	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
2	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
3	3	1	0.63	0.63	0.60	0.52	0.39	0.19	0.14	0.22
4	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
5	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
6	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
7	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
8	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
9	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
10	2	1	0.42	0.42	0.33	0.30	0.33	0.19	0.15	0.20
11	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
12	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
13	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
14	2	1	0.42	0.42	0.33	0.30	0.33	0.19	0.15	0.20
15	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
16	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
17	6	0	1.25	1.25	1.73	1.54	0.25	0.56	0.44	0.43
18	5	0	1.04	1.04	1.31	1.13	0.31	0.58	0.50	0.49
19	11	10	2.29	2.29	4.32	4.50	0.00	0.01	0.02	0.06
20	6	2	1.25	1.25	1.73	1.54	0.26	0.10	0.11	0.11
21	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
22	3	1	0.63	0.63	0.60	0.52	0.39	0.19	0.14	0.22
23	3	1	0.63	0.63	0.60	0.52	0.39	0.19	0.14	0.22
24	4	0	0.83	0.83	0.93	0.79	0.39	0.62	0.58	0.57
25	4	0	0.83	0.83	0.93	0.79	0.39	0.62	0.58	0.57
26	2	2	0.42	0.42	0.33	0.30	0.04	0.08	0.09	0.05
27	4	1	0.83	0.83	0.93	0.79	0.41	0.18	0.12	0.22
28	5	1	1.04	1.04	1.31	1.13	0.41	0.18	0.11	0.21
29	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
30	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
31	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
32	6	2	1.25	1.25	1.73	1.54	0.26	0.10	0.11	0.11

33	2	2	0.42	0.42	0.33	0.30	0.04	0.08	0.09	0.05
34	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
35	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
36	3	3	0.63	0.63	0.60	0.52	0.01	0.05	0.07	0.04
37	3	0	0.63	0.63	0.60	0.52	0.50	0.66	0.66	0.65
38	3	0	0.63	0.63	0.60	0.52	0.50	0.66	0.66	0.65
39	2	1	0.42	0.42	0.33	0.30	0.33	0.19	0.15	0.20
40	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
41	4	0	0.83	0.83	0.93	0.79	0.39	0.62	0.58	0.57
42	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
43	3	1	0.63	0.63	0.60	0.52	0.39	0.19	0.14	0.22
44	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
45	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
46	3	1	0.63	0.63	0.60	0.52	0.39	0.19	0.14	0.22
47	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
48	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
49	2	2	0.42	0.42	0.33	0.30	0.04	0.08	0.09	0.05
50	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
51	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
52	2	1	0.42	0.42	0.33	0.30	0.33	0.19	0.15	0.20
53	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
54	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
55	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
56	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
57	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
58	2	0	0.42	0.42	0.33	0.30	0.63	0.73	0.76	0.75
59	4	0	0.83	0.83	0.93	0.79	0.39	0.62	0.58	0.57
Total	163	34								

Notes:
BB = Beta Binomial model.
Bin = Binomial model.
IR = Incremental Risk model.
FH = Family History model.

References

Agresti, A., 2007. Introduction to Categorical Data Analysis, 2nd ed. Wiley–Interscience.

Association of American Railroads, 2017. Railroads and Chemicals. (Available from: https://www.aar.org/BackgroundPapers/Railroads%20and%20Chemicals.pdf).

Bagheri, M., Saccomanno, F.F., Chenouri, S., Fu, L.P., 2011. Reducing the threat of in-transit derailments involving dangerous goods through effective placement along the train consist. Accid. Anal. Prev. 43, 613–620.

Bagheri, M., Saccomanno, F.F., Chenouri, S., Fu, L.P., 2012. Modelling hazardous materials risks for different train make-up plans. Transp. Res. Part E: Logist. Transp. Rev. 48, 907–918.

Bagheri, M., Verma, M., Verter, V., 2014. Transport mode selection for toxic gases: rail or road? Risk Anal. 34 (1), 168–186.

Barkan, C.P.L., Saat, M.R., González, F., Treichel, T., 2013. Cooperative research in tank car safety design. TR News 286, 12–19.

Glickman, T.S., Erkut, E., Zschocke, M.S., 2007. The cost and risk impacts of rerouting railroad shipments of hazardous materials. Accid. Anal. Prev. 39 (5), 1015–1025.

Liu, X., Hong, Y., 2015. Modeling railroad tank car releases using a generalized binomial model. Accid. Anal. Prev. 84, 20–26.

Liu, X., Saat, M.R., Barkan, C.P.L., 2012. Analysis of causes of major train derailment and their effect on accident rates. Transp. Res. Rec. 2289, 154–163.

Liu, X., Saat, M.R., Barkan, C.P.L., 2014. Probability analysis of multiple-tank-car release incidents in railway hazardous materials transportation. J. Hazard. Mater. 276, 442–451.

Liu, X., 2016. Statistical causal analysis of freight-train derailments in the United States. J. Transp. Engineering Part A: Systems 143 (2), 04016007.

Liu, X., 2017. Optimizing rail defect inspection frequency to reduce the risk of hazardous materials transportation by rail. J. Loss Prev. Process Ind. 48, 151–161.

Nayak, P.R., Rosenfield, D.B., Hagopian, J.H., 1983. Event probabilities and impact zones for hazardous materials accidents on railroads, in: Report DOT/FRA/ORD-83/20, FRA, U.S. Department of Transportation.

Treichel, T.T., Hughes, J.P., Barkan, C.P.L., Sims, R.D., Philips, E.A., Saat, M.R., 2006. Safety Performance of Tank Cars in Accidents: Probability of Lading Loss, RSI-AAR Railroad Tank Car Safety Research and Test Project. Association of American Railroads, Washington D.C.

Yu, C., Zelterman, D., 2002. Sums of dependent Bernoulli random variables and disease clustering. Stat. Prob. Lett. 57, 363–373.