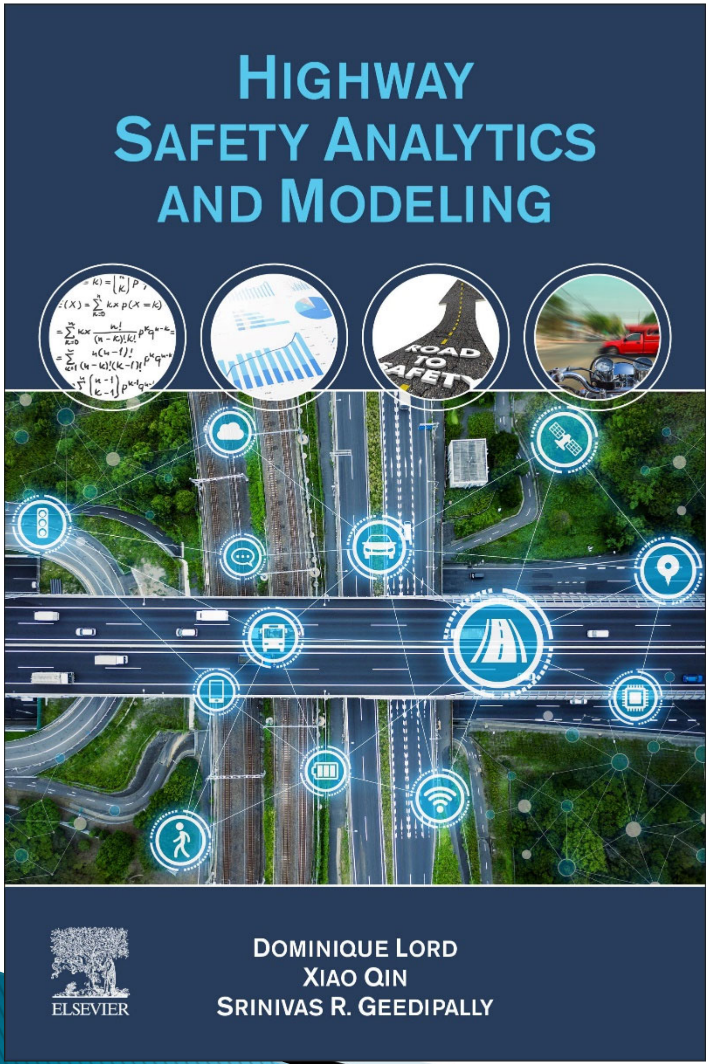
Crash-Frequency Models

Part 2

November 10, 2021

Instructor: Dominique Lord
Texas A&M University
d-lord@tamu.edu

Textbook



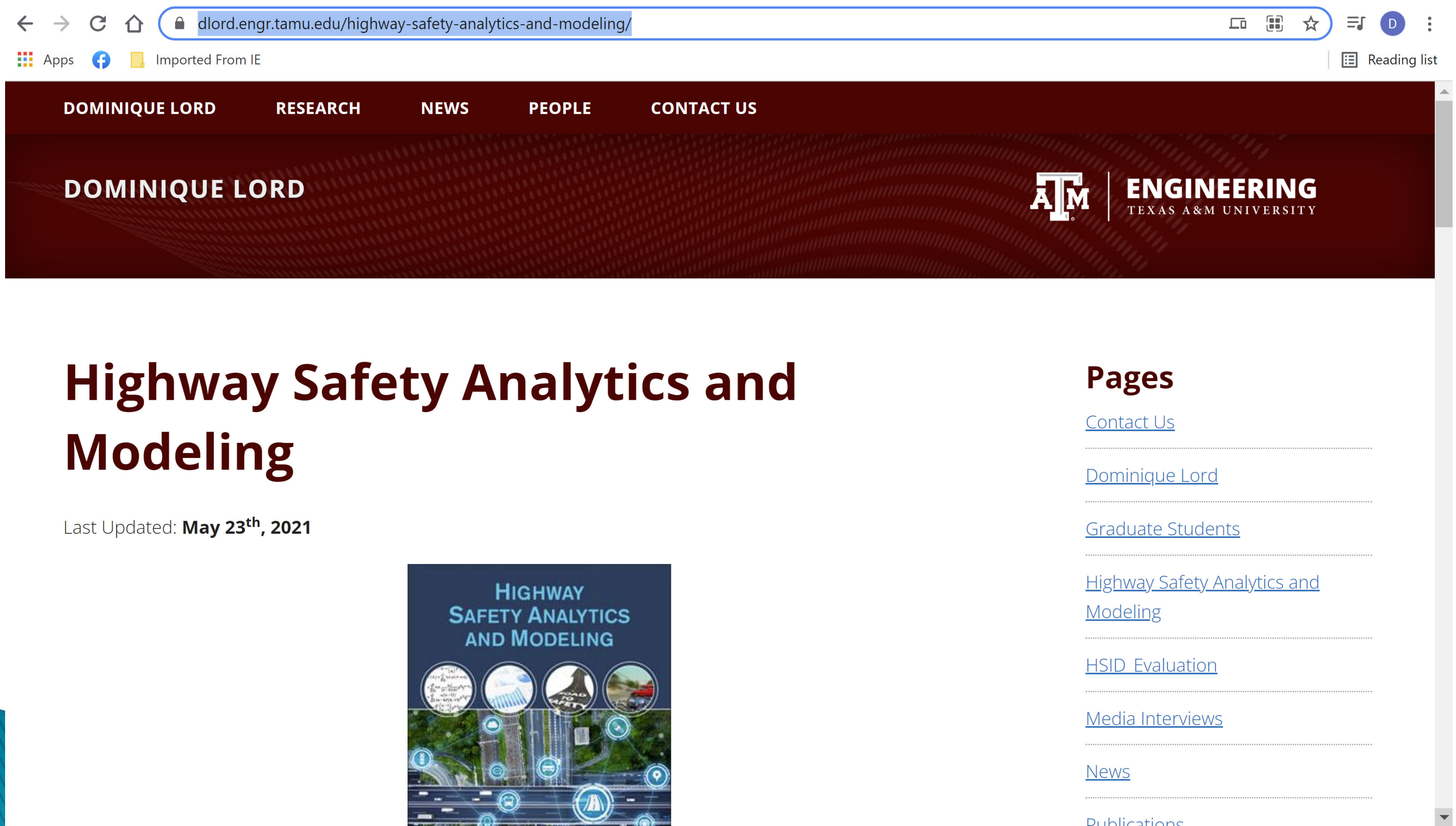
The material presented in this series of lectures are taken from this textbook and other sources based on lectures given by the authors.

The textbook is available on Amazon and the Elsevier website below among other places.

https://www.elsevier.com/books/highway-safety-analytics-and-modeling/lord/978-0-12-816818-9

Textbook

Datasets for examples and updates/corrections can be find in the following link:
https://dlord.engr.tamu.edu/highway-safety-analytics-and-modeling/



Crash-Frequency Models

- Basic Models
  - Poisson, NB, PLN
- Generalized count models for underdispersion
  - COM-Poisson
- Finite mixture and multivariate models
- Multi-distribution models
  - NB-L, NB-GE
- Models for better capturing unobserved heterogeneity
  - Random Effects, Random Parameters
- Semi- and nonparametric models
  - GAMs, Semi-parametric Poisson, NB-Dirichlet process
  - SVM, MLP (Neural Network), BNN

Model Selection

- Selection of model should be based on characteristics of data and study objectives
- Pragmatic approach: MLE (for simple) and Bayesian (for complex)
- Goodness-of-fit not the sole objective
- Goodness-of-logic is also important
- It is good to introduce new models/methods, but it should address a specific problem
- Although the NB model can suffer from methodological issues, it is still a very solid model that has been studied extensively

Outline

- Empirical Bayes (EB) Method
- Crash Variance and Variance Function
- Low Sample Mean and Small Sample Size
- Goodness of Logic
- Reducing Unobserved Heterogeneity
- Correlation between Severity Models
- Safety Performance of Automated Vehicles (New topic)

Empirical Bayes Method

(introduction – it will be discussed more as part of Chapter 7)

EB Method

The empirical Bayes (EB) method was initially developed as an approximation of the Bayes method, as the latter method requires a multidimensional integration of the total prior function of the Bayesian (bottom of equation), which was first described in Chapter 2:

$$f(\mu|y) = \frac{p(y|\mu)f(\mu)}{\int_a^b p(y|\mu)f(\mu)d\mu}$$

With the availability of very advanced computing power and the application of the Markov Chain Monte Carlo (MCMC), the EB method is basically no longer use in statistics and other fields. However, it remains very popular in highway safety.

The EB method was initially introduced by researchers from England (TRRL) in a 1981 paper (Abbess et al., 1981) and later refined by Ezra Hauer in the late 80s and early 90s.

EB Method

In highway safety, the EB method consists of using two sources of information: **one coming from the site investigated** and **one coming from a population of sites that are assumed to have the same basic characteristics** (e.g., 4-legged urban signalized intersections, 4-lane divided rural arterials). The EB estimate is calculated as follows (based on the NB distribution/model):

$$\mu_{i\_EB} = \gamma_i \mu_i + (1 - \gamma_i) y_i$$

$$\gamma_i = \frac{1}{1 + \frac{\mu_i}{\phi}}$$

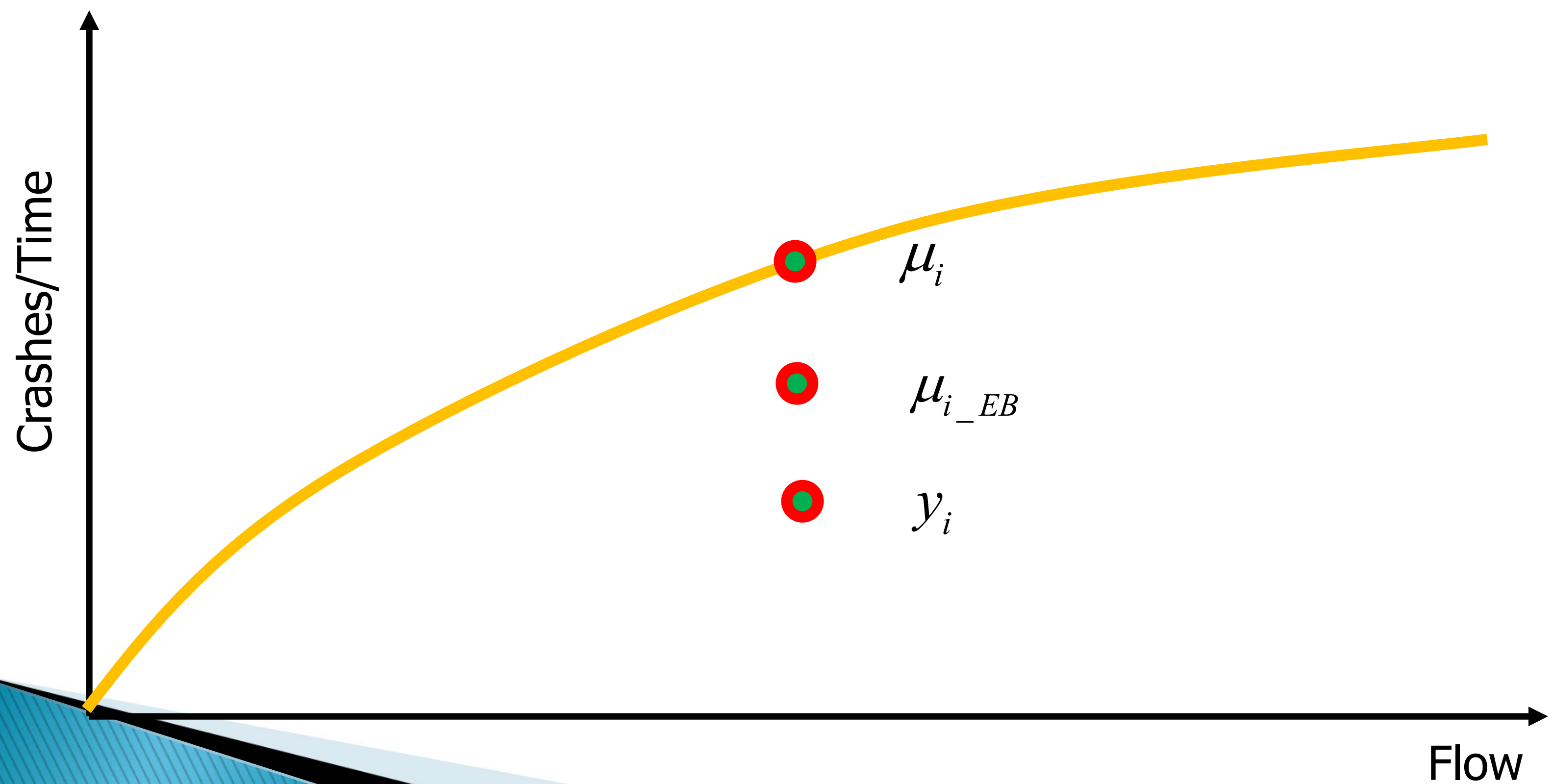
$$\mu_i = \exp\left(\mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i\right)$$

Gamma (1, $1/\phi$)

Note: In the context of modeling for estimating the safety of entities, the same sites/observations y_i are used for estimating μ_i. For statisticians, this is a violation of the Bayes rules (using data twice).

EB Method

In the context of safety estimation, the EB method is assumed to more accurately estimate the long-term mean of a given site. Recall that the simplified assumption states that crashes for a given site/observations follow a Poisson distribution (over time) where the mean is gamma distributed (or other distributions).



Crash Modeling and Variance Function

Variance Functions

Fixed: $$Var(y) = \mu + \alpha\mu^2$$

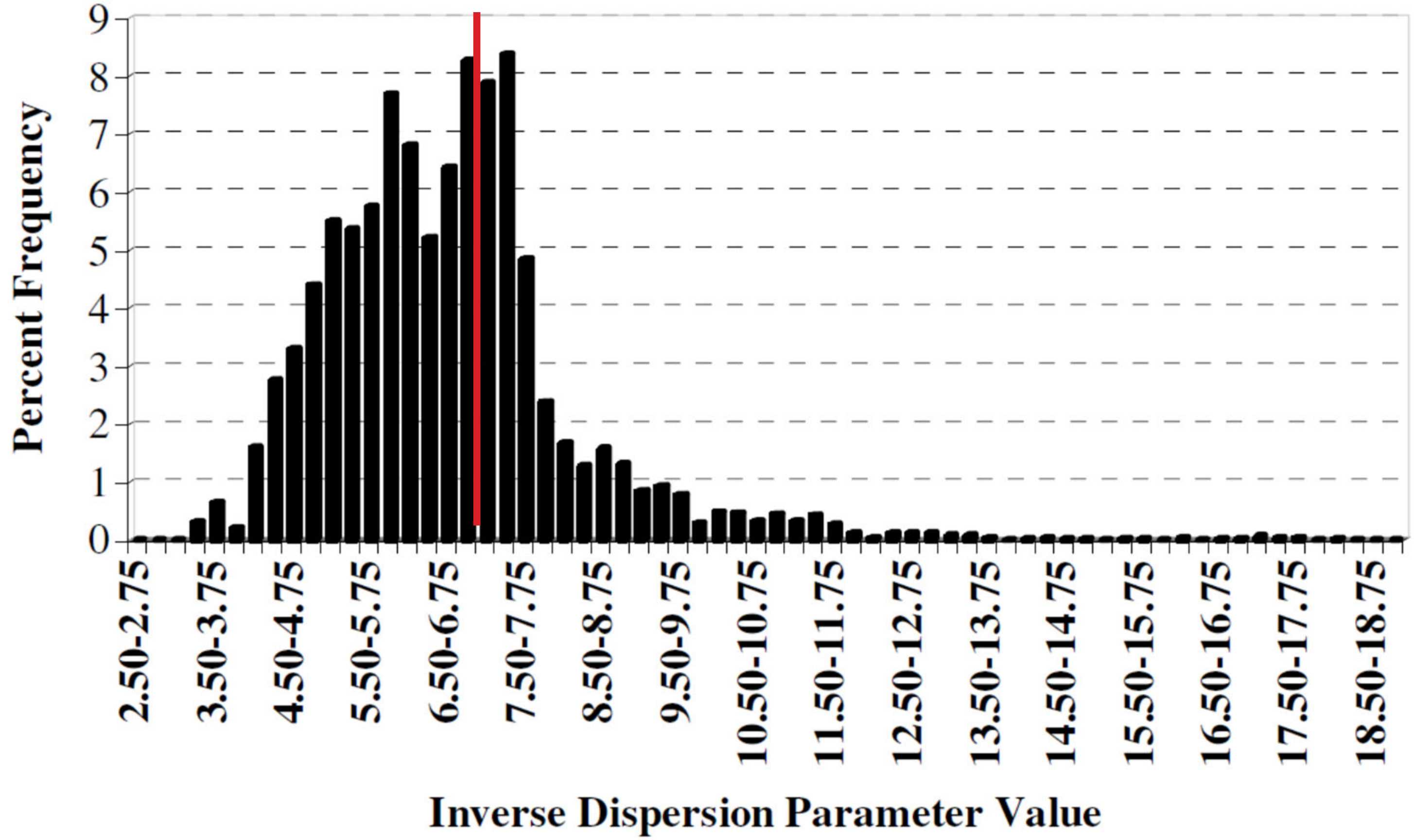
Function of segment length: $$\alpha_i = e^{\gamma_0} L_i^{\gamma_1}$$

$$\alpha_i = \frac{1}{e^{\gamma_0} L_i}$$

Function of covariates: $$\alpha_i = e^{\gamma_0 + \gamma_1 \times AADT_{maj,i} + \gamma_2 \times AADT_{min,i} + \gamma_3 \times AADT_{min},i / AADT_{maj,i}}$$

$$\alpha_i = \exp\left(\mathbf{z}_i'\gamma + \varpi_i\right)$$

Variance Functions



$$\phi_{it} = \exp(\eta_0 + \eta_1 F_{1,it} + \eta_2 F_{2,it} + \eta_3 F_{2,it} / F_{1,it})$$

2.2200	-0.02126	0.05452	-0.8783
(±0.180)	(±0.005)	(±0.009)	(±0.265)

Variance Functions

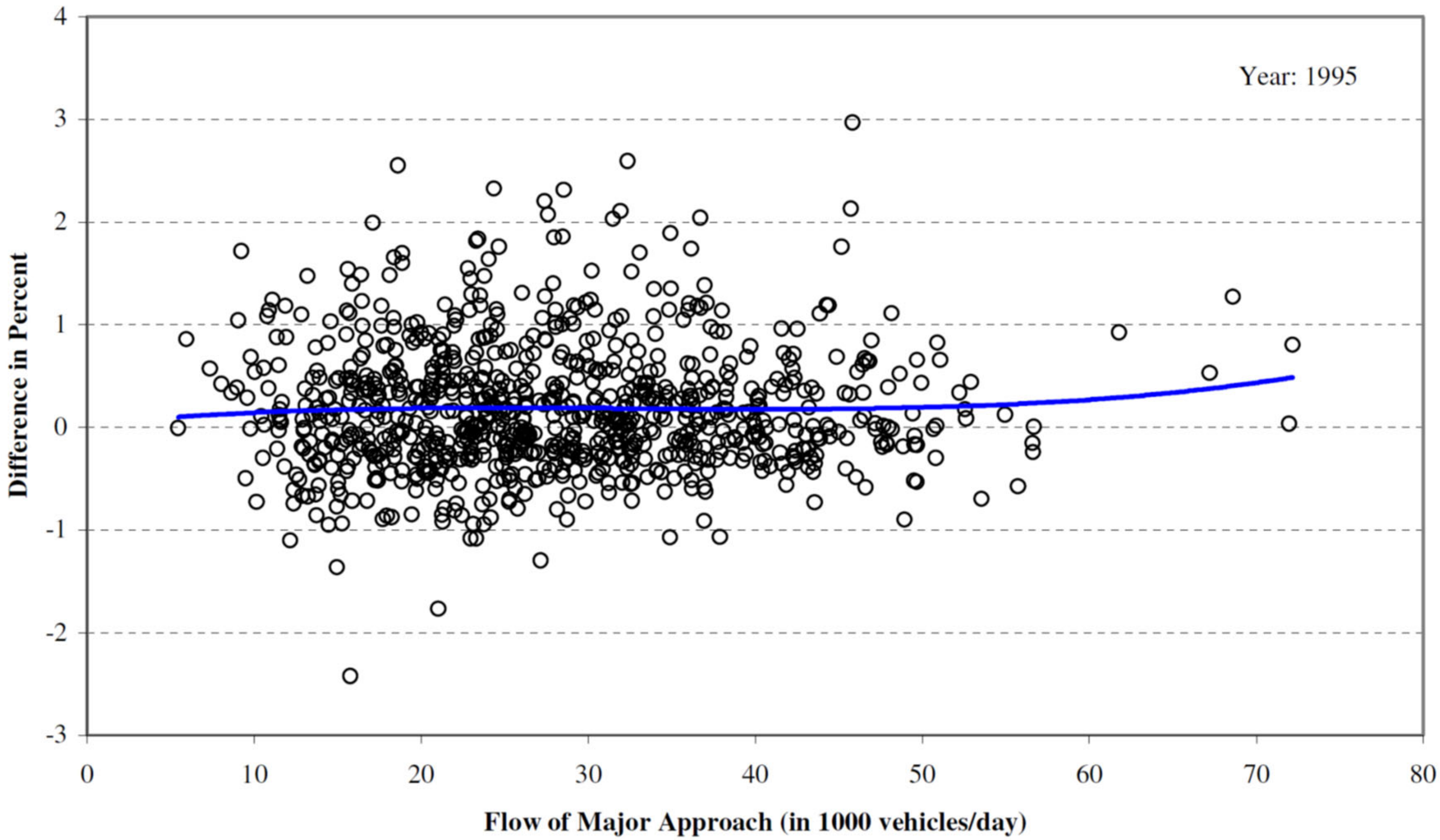


FIGURE 5. Differences in estimates of expected number of crashes for individual sites: EB versus Bayes, both with a fixed dispersion assumption.

Variance Functions

Many other models allow for a varying dispersion parameter. They include the COM-Poisson and Poisson-Inverse Gaussian among others.

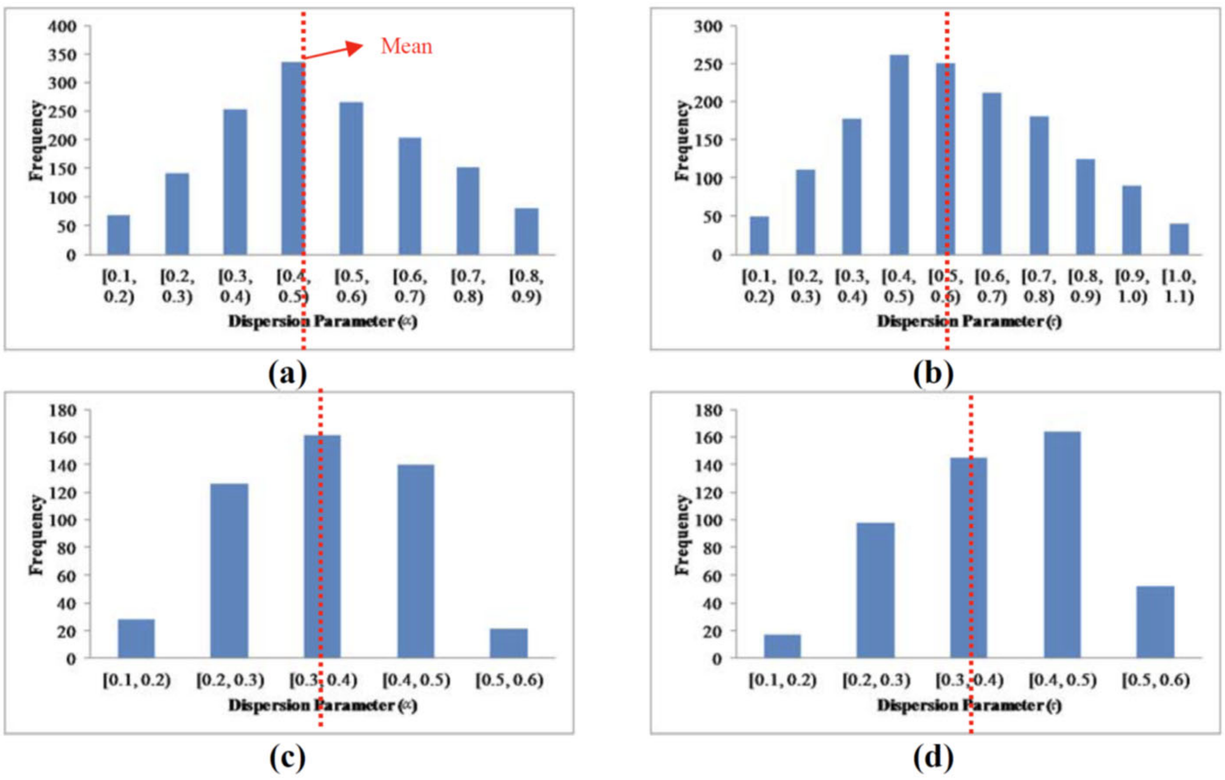


Figure 3 Distribution of Dispersion Parameter of NB and PIG Model for Texas Data ((a), (b)) and Washington Data ((c), (d)).

Low Sample Mean and Small Sample Size

LSM and SSS

Table 3. Simulation Results for $\lambda = 0.5$ (Fixed Mean)

Characteristics	$\phi = 1/2$				$\phi = 1$				$\phi = 2$			
	$\hat{\lambda}$	MM[1]	WR[2]	ML[3]	$\hat{\lambda}$	MM	WR	ML	$\hat{\lambda}$	MM	WR	ML
	n=50[†]				n=50				n=50			
Mean	0.54 (0.15)[‡]	0.70 (0.35)	0.71 (0.36)	0.67 (0.36)	0.51 (0.09)	3.18 (4.67)	3.25 (4.77)	2.74 (3.68)	0.52 (0.10)	6.04 (8.18)	6.17 (8.35)	5.72 (8.08)
Max	0.96	1.67	1.70	1.80	0.66	25.75	26.27	20.10	0.74	29.65	30.25	32.98
Min	0.32	0.20	0.21	0.27	0.34	0.52	0.53	0.47	0.30	0.91	0.92	1.21
	n=100				n=100				n=100			
Mean	0.47 (0.11)	0.74 (0.38)	0.75 (0.39)	0.67 (0.29)	0.52 (0.09)	1.45 (0.98)	1.47 (0.99)	1.28 (0.83)	0.53 (0.10)	3.90 (3.83)	3.94 (3.87)	3.67 (3.42)
Max	0.67	1.89	1.91	1.45	0.68	4.30	4.34	3.48	0.70	19.72	19.92	17.82
Min	0.26	0.23	0.23	0.29	0.32	0.36	0.36	0.33	0.37	1.35	1.36	1.25
	n=1000				n=1000				n=1000			
Mean	0.51 (0.03)	0.52 (0.07)	0.52 (0.07)	0.49 (0.06)	0.49 (0.03)	1.01 (0.21)	1.01 (0.21)	1.00 (0.20)	0.51 (0.02)	2.08 (0.47)	2.08 (0.47)	2.09 (0.47)
Max	0.56	0.78	0.78	0.75	0.54	1.53	1.53	1.51	0.56	3.00	3.00	3.10
Min	0.46	0.38	0.38	0.41	0.44	0.66	0.66	0.69	0.47	1.29	1.29	1.29

[†] Sample size, [‡] Standard deviation, [1] Method of moments (Estimator 1), [2] Weighted Regression (Estimator 2), [3] Maximum Likelihood (Estimator 3)

LSM and SSS

Table 8. Effects of an Unreliably Estimated Dispersion Parameter ($\lambda = 0.5$)

y	Freq*	‡ $\phi = 1$		$\phi = 2$			$\phi = 3$		
		γ	$\hat{\hat{\mu}}$	γ	$\hat{\hat{\mu}}$	†Diff	γ	$\hat{\hat{\mu}}$	†Diff
0	67	0.67	0.33	0.80	0.40	20.0%	0.86	0.46	28.6%
1	26	0.67	0.67	0.80	0.60	10.0%	0.86	0.57	14.3%
2	5	0.67	1.00	0.80	0.80	20.0%	0.86	0.71	28.6%
3	1	0.67	1.33	0.80	1.00	25.0%	0.86	0.86	35.7%
4	0	--	--	--	--	--	--	--	--
5	1	0.67	2.00	0.80	1.40	30.0%	0.86	1.14	42.9%
Total	100								
Ave	0.44								

†Relative Difference, ‡ Theoretical value used for the simulation, *Frequency or number of observations

Sample Size

TABLE 6.4 Recommended sample size (Lord, 2006).

Population sample mean	Minimum sample size
5.00	200
4.00	250
3.00	335
2.00	500
1.00	1000
0.75	1335
0.50	2000
0.25	4000

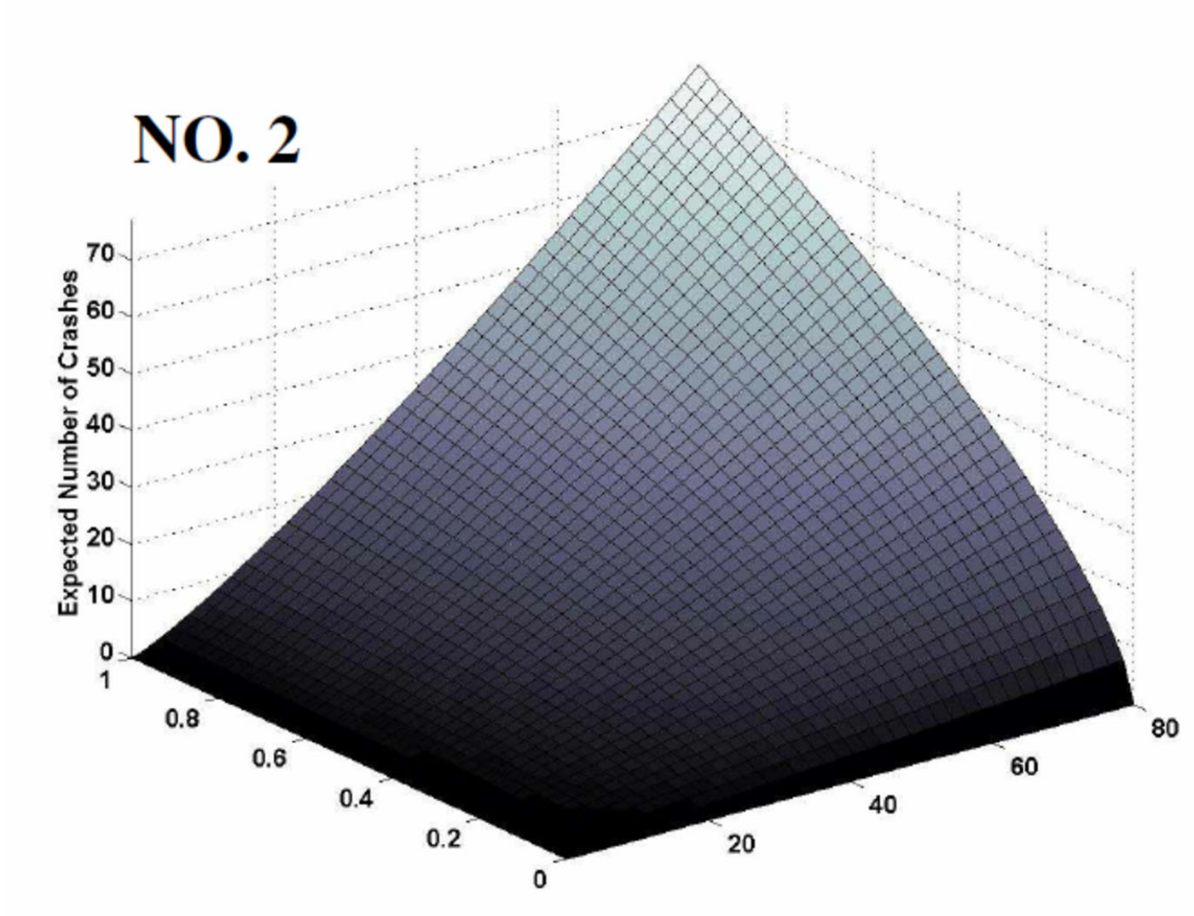
Sample Size

TABLE 6.5 Recommended minimum sample size for Bayesian Poisson-lognormal models (Miranda-Moreno et al., 2008).

Population sample mean	Minimum sample size
≥2.00	20
1.00	100
0.75	500
0.50	1000
0.25	3000

Goodness of Logic

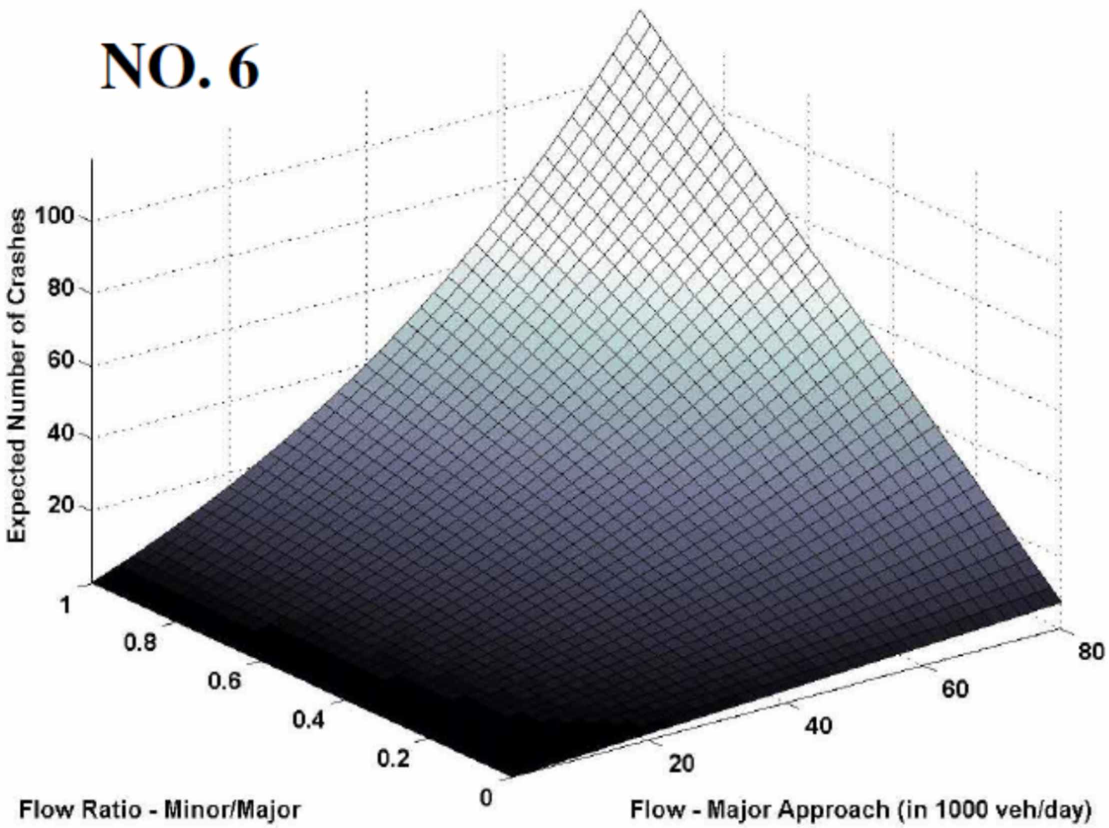
Goodness of Logic



$$\beta_{0,t} F_{1,it}^{\beta_1} F_{2,it}^{\beta_2}$$

If either F_1 or F_2 = 0, no crash can occur, but if no F_2, vehicles on major road can still turn and hit another vehicle on major road. (crash risk not zero)

Goodness of Logic



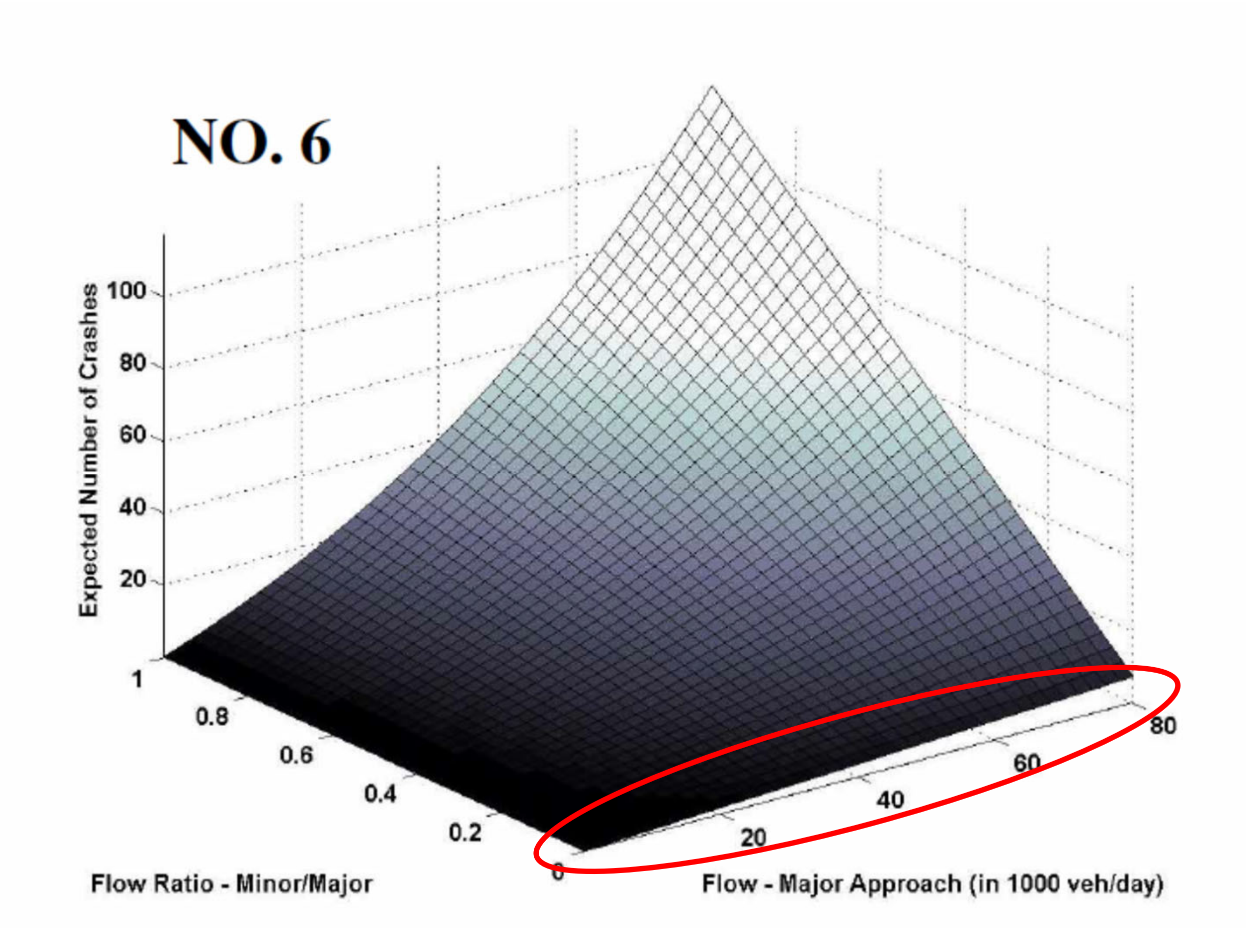
$$F_{1,it}\lambda_{1,it} + F_{2,it}\lambda_{2,it}$$

where

$$\lambda_{1,it} = \exp(\beta_{0,t} + \beta_1 F_{2,it})$$

$$\lambda_{2,it} = \exp(\beta_{0,t}^{*} + \beta_2 F_{1,it})$$

Goodness of Logic

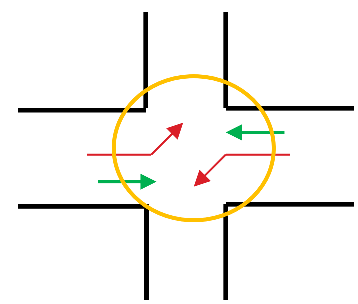


$$F_{1,it}\lambda_{1,it} + F_{2,it}\lambda_{2,it}$$

where

$$\lambda_{1,it} = \exp(\beta_{0,t} + \beta_1 F_{2,it})$$

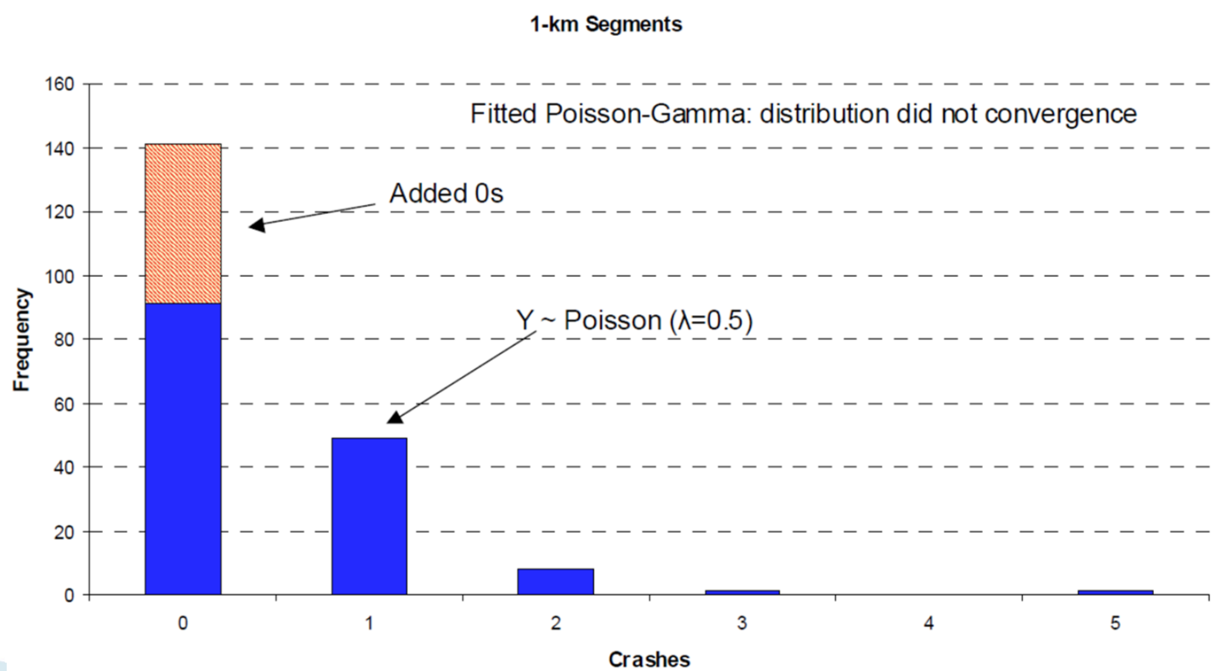
$$\lambda_{2,it} = \exp(\beta^*_{0,t} + \beta_2 F_{1,it})$$



Goodness of Logic

$$P(n) = \alpha + (1-\alpha)e^{-\lambda}; \qquad n = 0$$

$$P(n) = (1-\alpha)\frac{e^{-\lambda}\lambda^{n}}{n!}; \qquad n \geq 1$$



Goodness of Logic

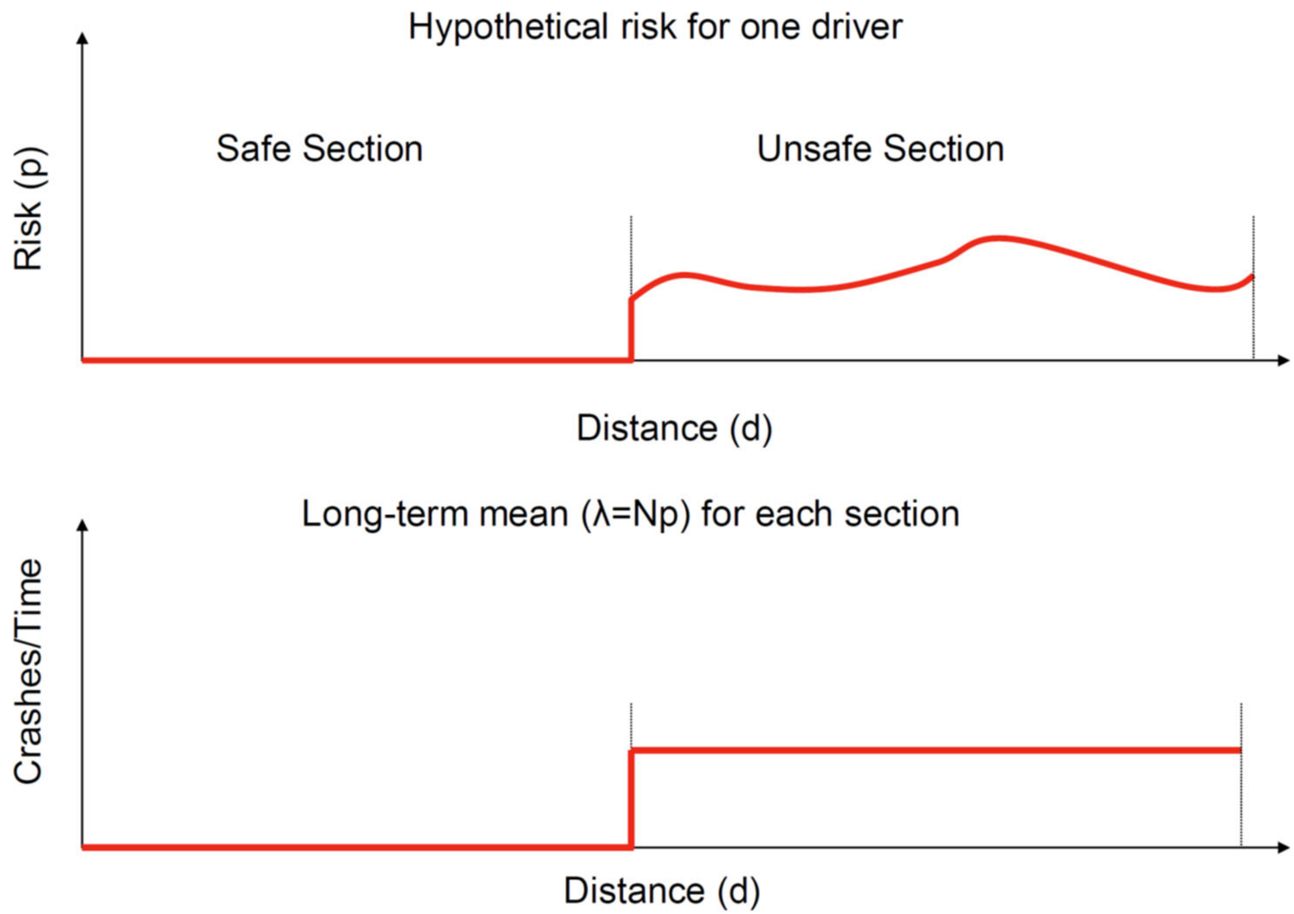


Figure 1. Crash Risk and Long-Term Mean of Two Hypothetical Highway Adjacent Sections

Goodness of Logic

Variable	Estimate Coefficients	*t*-Statistic
Zero-inflated State		
Constant	−10.731	−5.84
Environmental characteristics		
Visibility (miles)	0.959	1.78
Wet road surface indicator (1 if the road surface is wet, 0 otherwise)	−1.663	−3.64
Chemically Wet road surface indicator (1 if the road surface is chemically wet, 0 otherwise)	−1.864	−5.04
Traffic characteristics		
Hourly traffic volume (in 1000 vehicles per hour)	−0.611	−9.06
Truck percentage (%)	0.439	5.95
Temporal characteristics		
Night indicator (1 if the time period is at night, 0 otherwise)	0.352	1.94
Road characteristics		
Segment length (miles)	0.755	4.60
Number of lanes	1.917	6.10
Good pavement condition indicator (1 if the pavement condition is good, 0 otherwise)	0.680	2.63

Goodness of Logic

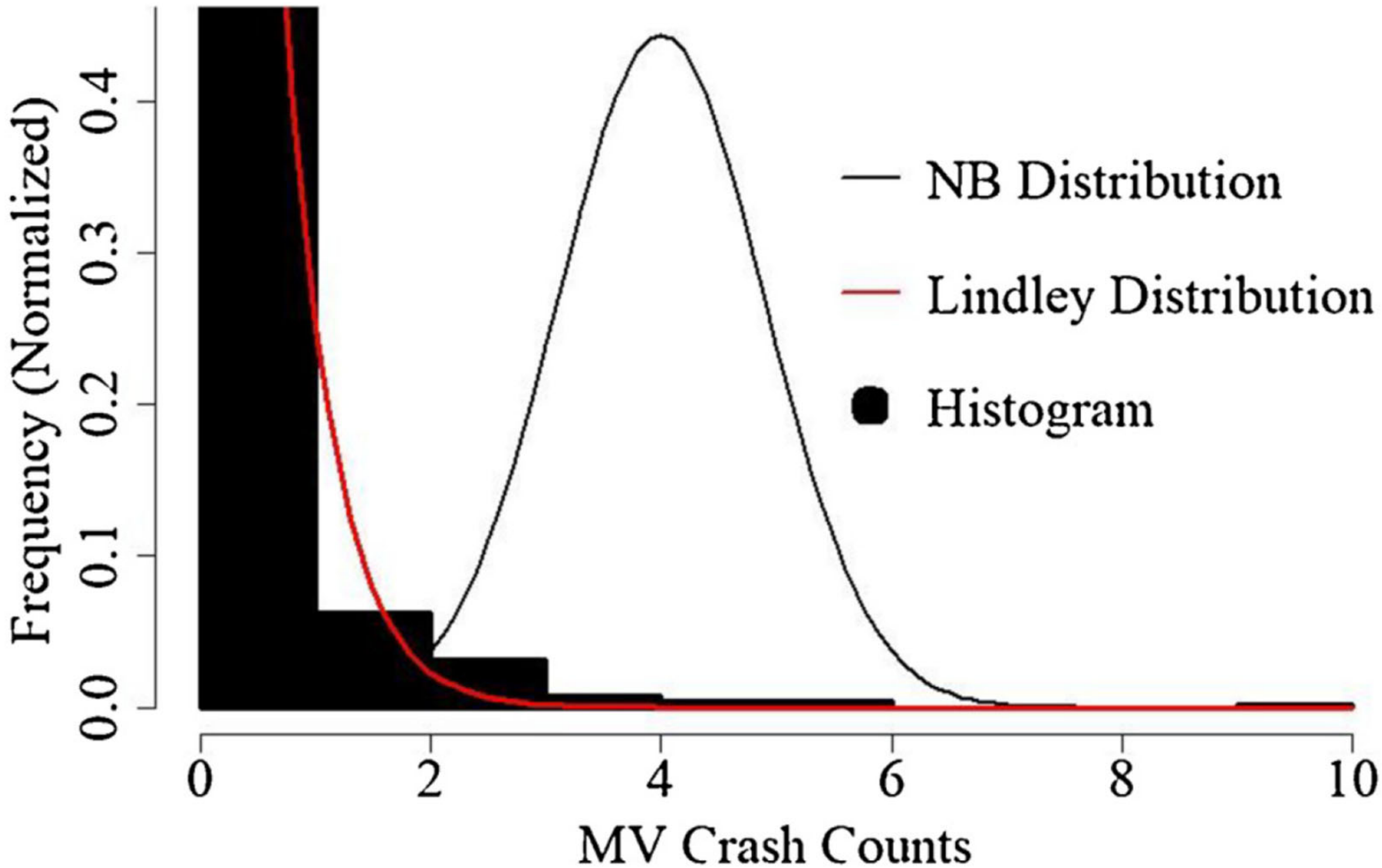
Negative Binomial State		
Constant	−10.673	−9.31
Environmental characteristics		
Cross wind speed (mph)	−0.013	−1.75
Wet road surface indicator (1 if the road surface is wet, 0 otherwise)	−0.529	−3.70
Traffic characteristics		
Low speed limit (1 if the speed limit is less than 60 mph, 0 otherwise)	0.387	1.83
Difference between speed limit and current traffic speed (speed limit minus traffic speed)	0.081	29.75
Truck percentage (%)	0.107	2.69
Temporal characteristics		
Sunset indicator (1 if the time period is during sunset, 0 otherwise)	−0.200	−1.88
November indicator (1 if the time period is in November, 0 otherwise)	0.292	3.20
4 am–5 am indicator (1 if the time period is between 4 am to 5 am, 0 otherwise)	−0.608	−1.96
Road characteristics		
Number of merging ramps per lane per mile	−1.072	−2.65
Segment length (miles)	0.786	5.44
Number of lanes	0.849	3.69
Curvature (degree)	0.406	3.07
Long Remaining service life of rutting indicator (1 if the value of ruti is higher than 99, 0 otherwise)	0.546	2.88
α	1.818	3.57
σ_i (site-specific)	0.484	7.54
Vuong statistic	4.48	
−2 Log Likelihood	15,145	

Goodness of Logic

TABLE 3. Posterior Means, Standard Deviations, 95% Credible Intervals for NB Mixtures

NB Mixtures		w	$Ln(\beta_0)$	β_1	β_2	ϕ
Single Component	Estimate (Std. dev.) (2.5%) (97.5%)	1.0	-10.2300 (0.4659) (-11.1707) (-9.3241)	0.6190 (0.0459) (0.5296) (0.7118)	0.6854 (0.0216) (0.6428) (0.7273)	7.0894 (0.6156) (5.9590) (8.3760)
Two Component	Component 1 Estimate (Std. dev.) (2.5%) (97.5%)	0.430 (0.153) (0.150) (0.731)	-10.9407 (1.3641) (-13.8766) (-8.3865)	0.8588 (0.1595) (0.5991) (1.2297)	0.5056 (0.0812) (0.3199) (0.6384)	9.3692 (1.6220) (6.8739) (12.9768)
	Component 2 Estimate (Std. dev.) (2.5%) (97.5%)	0.570 (0.153) (0.268) (0.849)	-9.7842 (1.0447) (-11.8434) (-7.6601)	0.3987 (0.1289) (0.1116) (0.6181)	0.8703 (0.0782) (0.7445) (1.0497)	8.2437 (1.3502) (6.0746) (11.2873)

Goodness of Logic



(a) Resemblance of Lindley distribution to the frequency distribution of observed multi-vehicle crash counts

Reducing Unobserved Heterogeneity

Reducing Unobserved Heterogeneity

- Important objective, but should not be the sole objective
- Two different philosophies:
  - Reduce it via the parameters (Random Parameters)
  - Reduce it via the error (multi-distribution, such as the NB-L)
- Do both at the same time (RPNB-L and other recent expansions)?

Reducing Unobserved Heterogeneity

Table 4. Modeling Results for the Indiana Data.

Variable	NB		NB-L		ZINB†	
	Value	Std. dev	Value	Std. dev	Value	Std. dev
INTERCEPT (β_0)	-4.779	0.979	-3.739	1.115	-8.3381	1.126
Ln(ADT) (β_1)	0.7219	0.091	0.630	0.106	1.0845	0.105
FRICTION (β_2)	-0.02774	0.008	-0.02746	0.0111	-0.0205	0.008
PAVEMENT (β_3)	0.4613	0.135	0.4327	0.217	0.2306	0.151
MW (β_4)	-0.00497	0.001	-0.00616	0.002	-0.0023	0.002
BARRIER (β_5)	-3.195	0.234	-3.238	0.326	-1.5095	0.389
RUMBLE (β_6)	-0.4047	0.131	-0.3976	0.213	-0.511	0.151
$\alpha = 1/\phi$	0.934	0.118	0.238	0.083	0.375	0.056
DIC[1]	1900		1701		1850‡	
MAD[2]	6.91		6.89		8.04	
MSPE[3]	206.76		195.54		268.01	
Pearson χ^2	1174		978		851	
MCPD[4]	454		261		778	

[1] Deviance Information Criterion; [2] Mean Absolute Deviance (Oh et al, 2003); [3] Mean Squared Predictive Error (Oh et al, 2003); [4] Maximum Cumulative Residual Plot Deviation (Geedipally et al, 2010).
†Estimated using the MLE and the inflated parameters are not presented here; ‡AIC

Reducing Unobserved Heterogeneity

Parameters	NB		RPNB		NB-L		RPNB-L	
	Value	Std. Dev.	Value	Std. Dev.	Value	Std. Dev.	Value	Std. Dev.
Parameter Mean								
Intercept	-4.449	0.067	-5.486	0.035	-3.947	0.162	-4.443	0.206
Log(ADT)	0.689	0.133	0.816	31.750	0.651	0.145	0.717	0.231
Friction	-0.027	0.011	-0.029	0.133	-0.027	0.012	-0.032	0.015
Pavement	0.422	0.189	0.588	0.012	0.445	0.210	0.605	0.281
Median Width	-0.005	0.002	-0.012	0.240	-0.006	0.002	-0.012	0.004
Barrier	-3.031	0.308	-6.614	0.003	-3.282	0.338	-6.152	0.898
Rumble	-0.405	0.186	-0.288	0.437	-0.404	0.207	-0.329	0.260
$\alpha = 1/\phi$	0.950	0.122	0.137	0.035	0.239	0.083	0.128	0.028
θ					1.464	0.180	1.414	0.173
Std. Deviation of Random Parameters								
Log(ADT)			0.302	0.172			0.232	0.137
Friction			0.057	0.011			0.056	0.011
Pavement			0.326	0.216			0.291	0.200
Median Width			0.028	0.003			0.028	0.003
Barrier			2.390	0.399			1.925	0.709
Rumble			0.379	0.242			0.310	0.183
Model Performance								
Dbar	1891.93		1481.09		1585.93		1422.70	
Dhat	1883.01		1296.86		1469.51		1276.00	
pD	8.92		184.22		116.41		146.30	
DIC	1900.84		~~1665.31[†]~~ 1736		1702.34		1569.00	
MAD[§]	6.92		6.90		6.88		6.71	

Note: [†] With the MLE RPNB, only three variables (logarithm of ADT, presence of median barrier and interior rumble strips) were found to be random. This increased the Deviance Information Criterion or DIC to 1736.

Reducing Unobserved Heterogeneity

Table 2
Posterior estimates of model parameters for multi-vehicle crashes along rural mountainous highways.

Variables	RPNB		RPNB-L		RPNB-GE	
	Mean (Std. Dev[a])	[95% BCI[b]]	Mean (Std. Dev.)	[95% BCI]	Mean (Std. Dev)	[95% BCI]
Constant	−7.634 (1.886)	[-11.690,-3.606]	−9.098 (1.610)	[-13.100,-6.273]	−8.025 (1.617)	[-10.630,-4.259]
Exposure variable						
Log (ADT x segment length)	0.473 (0.135)	[0.187,0.757]	0.517 (0.116)	[0.296,0.767]	0.511 (0.114)	[0.251,0.710]
Real-time weather conditions						
Heavy rainfall indicator at time of crash *(1 if 1-hour amount of rainfall is greater than 5.08 mm, 0 otherwise)*	0.860 (0.408)	[0.069,1.732]	0.897 (0.403)	[0.113,1.714]	0.936 (0.419)	[0.071,1.746]
Longitudinal grades						
Combination of horizontal and vertical alignment indicator; Category 4: *(1 if more than 50% of a segment has horizontal curve and absolute gradient > 4%, 0 otherwise)*[*]	−0.094 (0.217)	[-0.538,0.307]	−0.018 (0.232)	[-0.474,0.443]	0.010 (0.194)	[-0.419,0.344]
Standard deviation of distribution	0.352 (0.244)	[0.022,0.954]	0.310 (0.223)	[0.031,0.860]	0.279 (0.223)	[0.000,0.800]
Cross-sectional elements						
Presence of a passing lane *(1 if there is a passing lane along the segment, 0 otherwise)*	−1.164 (0.466)	[-2.089,-0.295]	−1.104 (0.476)	[-2.230,-0.246]	−1.139 (0.466)	[-2.094,-0.270]
Roadway and roadside features						
Number of minor junctions	0.212 (0.060)	[0.094,0.329]	0.213 (0.059)	[0.097,0.329]	0.216 (0.060)	[0.099,0.335]
Presence of road delineation *(1 if there are road delineations such as guide posts and chevron signs along the segment, 0 otherwise)*[*]	−0.209 (0.230)	[-0.675,0.221]	−0.237 (0.231)	[-0.710,0.191]	−0.241 (0.265)	[-0.763,0.260]
Standard deviation of distribution	0.483 (0.346)	[0.049,1.409]	0.522 (0.332)	[0.076,1.335]	0.565 (0.381)	[0.082,1.587]
Dispersion parameter	1.901 (0.661)	[0.998,3.576]	1.911 (0.674)	[0.996,3.589]	1.884 (0.645)	[0.999,3.454]
Lindley parameter (Theta)	–	–	0.862 (0.770)	[0.080,2.979]		
Parameter of GE Distribution (a)					2.031 (0.569)	[1.059,2.956]
Parameter of GE Distribution (b)					1.498 (0.289)	[1.023,1.972]
Number of parameters	7		7		7	
LL	425.650		425.550		425.700	
DIC	885.100		752.800		870.500	
MAD	0.541		0.539		0.541	
MSPE	0.690		0.672		0.674	
MSE	0.690		0.687		0.690	

[*] random parameter.
[a] Std. Dev: Standard deviation of posterior estimates.
[b] BCI: Bayesian Credible Interval.

Reducing Unobserved Heterogeneity

Table 3
Analysis of FSI Crashes using Bayesian Inference and MCMC Simulation.

Response variable: FSI Crashes	Random Parameters Negative Binomial			Random Parameters NB with Heterogeneity in Means			Random Parameters NB-Lindley			Random Parameters NB with Heterogeneity in Means - Lindley		
Explanatory Variables	Mean	Std. Dev	MC error	Mean	Std. Dev	MC error	Mean	Std. Dev	MC error	Mean	Std. Dev	MC error
Constant	−11.96	0.82	0.03	−12.80	1.32	0.05	−14.03	1.26	0.05	−14.43	1.84	0.07
Logarithm of AADT	0.083	0.06	0.02	0.15	0.09	0.03	0.21	0.07	0.03	0.14	0.17	0.07
Logarithm of AADT (Std. Dev.)	0.028	0.02	0.01	0.03	0.02	0.01	0.03	0.02	0.01	0.03	0.02	0.01
Logarithm of MCV volume	1.03	0.12	0.04	0.98	0.12	0.04	0.97	0.12	0.04	1.02	0.17	0.06
Urban single carriageway	1.58	0.61	0.02	2.13	1.05	0.04	2.13	0.65	0.02	1.99	0.88	0.03
Rural single carriageway with high-speed limit	2.82	0.49	0.02	3.43	1.01	0.04	3.48	0.53	0.02	3.29	0.82	0.03
Rural single carriageway with high-speed limit (Std. Dev.)	0.20	0.18	0.01	0.21	0.16	0.01	0.17	0.14	0.01	0.20	0.17	0.01
Rural single carriageway with medium-speed limit	1.59	0.97	0.02	2.21	1.29	0.04	2.11	1.01	0.02	2.04	1.17	0.03
Curve longer than 50% of the segment length	−0.53	0.56	0.02	−0.38	0.64	0.02	−0.61	0.52	0.02	−0.47	0.69	0.02
Curve longer than 50% of the segment length (Std. Dev.)	0.68	0.68	0.02	0.62	0.63	0.02	0.69	0.71	0.02	0.79	0.74	0.02
	Heterogeneity in Means of Random Parameters											
Curve longer than 50% of the segment length: Roll Terrain	–	–	–	−0.56	0.91	0.02	–	–	–	−0.54	0.96	0.02
Dispersion parameter (α)	2.66	1.29	0.01	2.66	1.28	0.01	3.06	1.39	0.01	3.01	1.39	0.01
Average Log Likelihood	−283.12			−283.30			−269.69			−270.37		
DIC	603.09			604.75			568.77			589.56		
MAD	0.089	0.01	0.00	0.088	0.01	0.00	0.085	0.01	0.00	0.086	0.01	0.00
MSPE	0.121	0.03	0.00	0.122	0.04	0.00	0.123	0.04	0.00	0.125	0.05	0.00
theta	–	–	–	–	–	–	1.29	1.05	0.04	0.64	0.78	0.03

Reducing Unobserved Heterogeneity

Response variable: FSI Crashes	Random Parameters Poisson			Random Parameters Poisson with Heterogeneity in Means			Random Parameters Poisson-Lindley			Random Parameters Poisson with Heterogeneity in Means - Lindley		
Explanatory Variables	Mean	Std. Dev	MC error	Mean	Std. Dev	MC error	Mean	Std. Dev	MC error	Mean	Std. Dev	MC error
Constant	−13.08	1.22	0.05	−13.26	1.37	0.05	−13.24	1.58	0.06	−13.20	1.23	0.05
Logarithmic of AADT	0.16	0.13	0.05	0.21	0.09	0.04	0.07	0.11	0.04	0.31	0.13	0.05
Logarithmic of AADT (Std. Dev.)	0.03	0.03	0.01	0.03	0.02	0.01	0.04	0.03	0.01	0.02	0.02	0.01
Logarithmic of MCV volume	0.98	0.15	0.05	0.93	0.12	0.04	1.04	0.13	0.05	0.89	0.14	0.05
Urban single carriageway	2.32	0.98	0.04	2.28	0.92	0.03	1.76	0.76	0.03	1.78	0.63	0.02
Rural single carriageway with high-speed limit	3.58	0.91	0.04	3.62	0.90	0.04	2.99	0.71	0.03	3.20	0.51	0.02
Rural single carriageway with high-speed limit (Std. Dev.)	0.31	0.21	0.01	0.24	0.19	0.01	0.19	0.17	0.01	0.14	0.13	0.00
Rural single carriageway with medium-speed limit	2.37	1.24	0.04	2.36	1.18	0.03	1.80	1.07	0.03	1.81	1.00	0.02
Curve longer than 50% of the segment length	−0.65	0.53	0.02	−0.51	0.62	0.02	−0.64	0.48	0.02	−0.56	0.62	0.02
Curve longer than 50% of the segment length (Std. Dev.)	0.70	0.69	0.02	0.71	0.70	0.02	0.66	0.70	0.02	0.67	0.68	0.02
	Heterogeneity in Means of Random Parameters											
Curve longer than 50% of the segment length: Roll Terrain	–	–	–	−0.49	0.89	0.02	–	–	–	−0.47	0.93	0.02
Average Log Likelihood	−274.50			−276.00			−262.74			−253.56		
DIC	594.73			594.75			586.70			550.03		
MAD	0.084	0.01	0.00	0.085	0.01	0.00	0.082	0.01	0.00	0.079	0.01	0.00
MSPE	0.106	0.01	0.00	0.107	0.01	0.00	0.104	0.01	0.00	0.103	0.01	0.00
theta	–	–	–	–	–	–	0.90	0.77	0.03	1.55	1.21	0.05

Correlation between Crash Severities

Crash Severity Correlation

	Univariate	Multivariate
Constant	**-9.9596** (0.6670)	**-10.1806** (0.3065)
Lighting	**0.4203** (0.1051)	**0.3544** (0.0465)
Painted Left Turn	**-0.2159** (0.1127)	**-0.2326** (0.0420)
Curb Med Left Turn	-0.1494 (0.1482)	**-0.1836** (0.0611)
Rhgt Trn Channel	0.0715 (0.1263)	**0.1864** (0.0525)
ML Lanes	0.1257 (0.0723)	**0.1041** (0.0373)
Mountain	**0.5337** (0.1347)	**0.5352** (0.0533)
Rolling	0.1260 (0.1046)	**0.1403** (0.0437)
Logmaj	**0.9777** (0.0717)	**1.0593** (0.0315)
Logmin	**0.2493** (0.0333)	**0.2193** (0.0132)

Estimating the Safety Performance of Automated Vehicles

Safety Performance of AVs

- A lot of research is now devoted to AVs
- Data related to crashes or near misses is very limited
  - In California, companies need to report all crashes involving AVs (within two weeks)
  - NHTSA is implementing such rule at the national level in the US
- This limited data availability causes important methodological challenges
  - Simulation work and risk probabilities (failure) when they are available
- Need for methods to estimate their safety

Safety Performance of AVs

Sohrabi et al. (2021) have proposed using a duration model/hazard function for such evaluation. The goal consists of estimating time to crash given the available exposure (aggregated). The approach was also used for comparing it with human-driven vehicles (*conventional vehicles* or CV).

The probability of survival (being crash-free) beyond x miles is shown as:

$$S(x) = \Pr(X > x) = \int_x^{\infty} f(x)\mathrm{dx}$$

The cumulative distribution function of number of miles between crashes can then be written as:

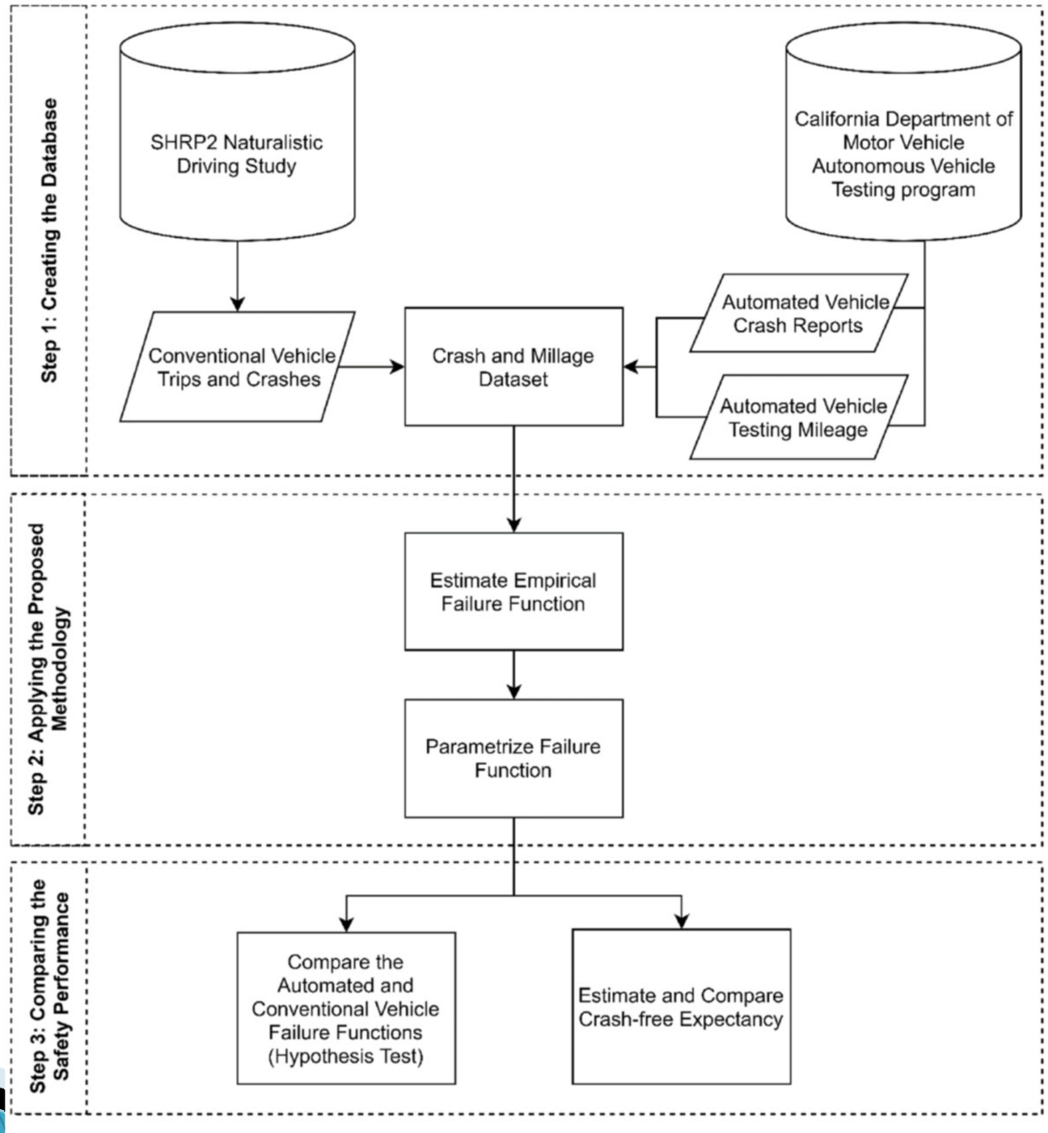
$$F(x) = 1 - S(x)$$

The first derivative of this cumulative distribution with respect to distance gives the density function $f(x) = dF(x)/dx$. With this, the instantaneous rate of failure (crashes) is represented by the hazard function, $h(x)$:

$$h(x) = f(x)/[1 - F(x)]$$

Sohrabi, S., D. Lord, B. Dadashova, F. Mannering (2021)Towards the assessment of automated-vehicle safety with duration modeling. Paper submitted for publication.

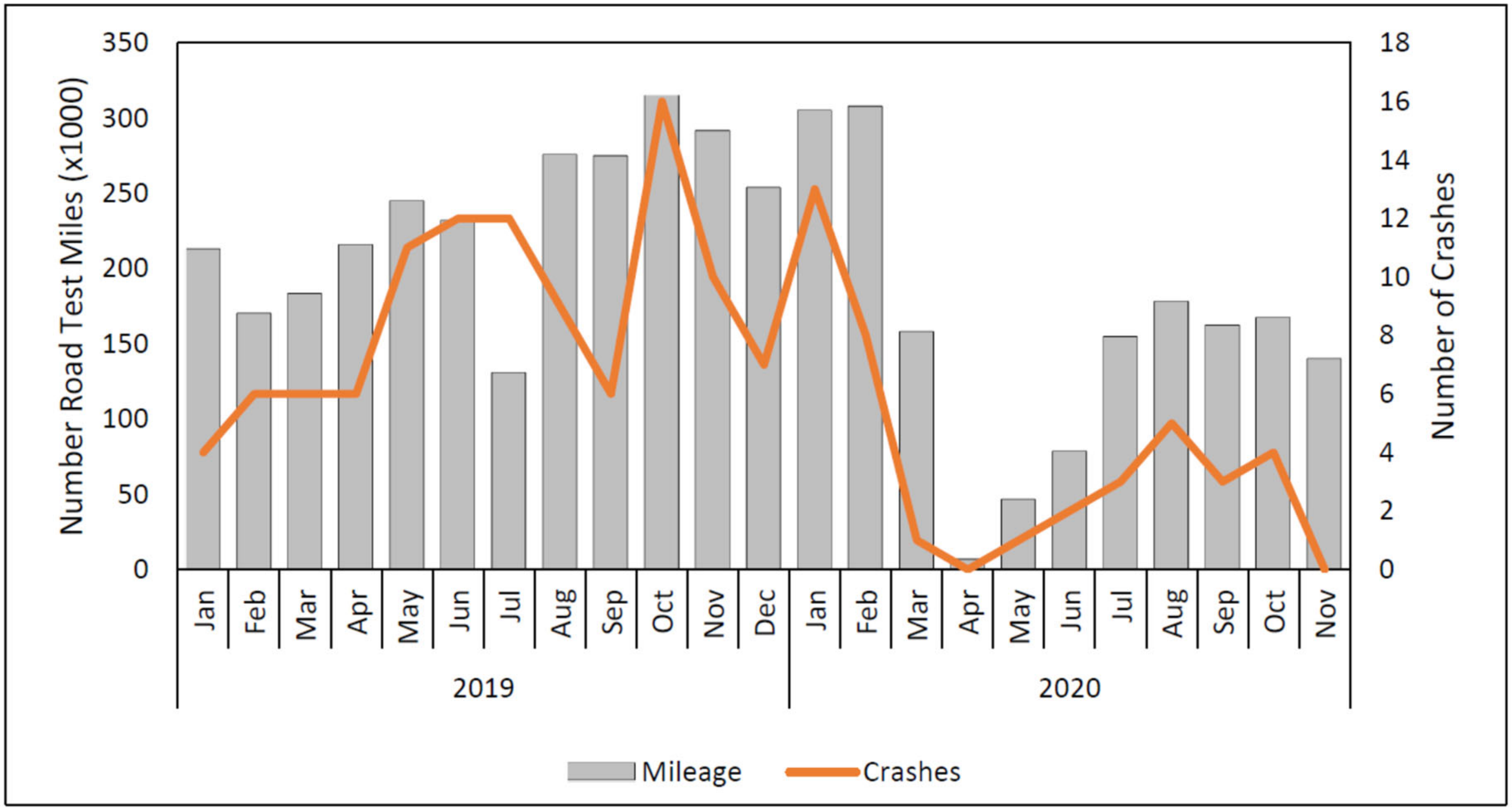
Safety Performance of AVs



Safety Performance of AVs

Descriptive Statistics	Conventional-vehicle crash data	Automated-vehicle crash data
Number of crashes	130	105
Number of miles driven (million miles)	2,849,850	2,849,850
Rate of crashes (per million miles)	45.6	36.8
Mean miles-to-crash	21,634	27,399
Minimum miles-to-crash	12	4,212
Maximum miles-to-crash	112,975	134,023
Median miles-to-crash	12,679	15,767

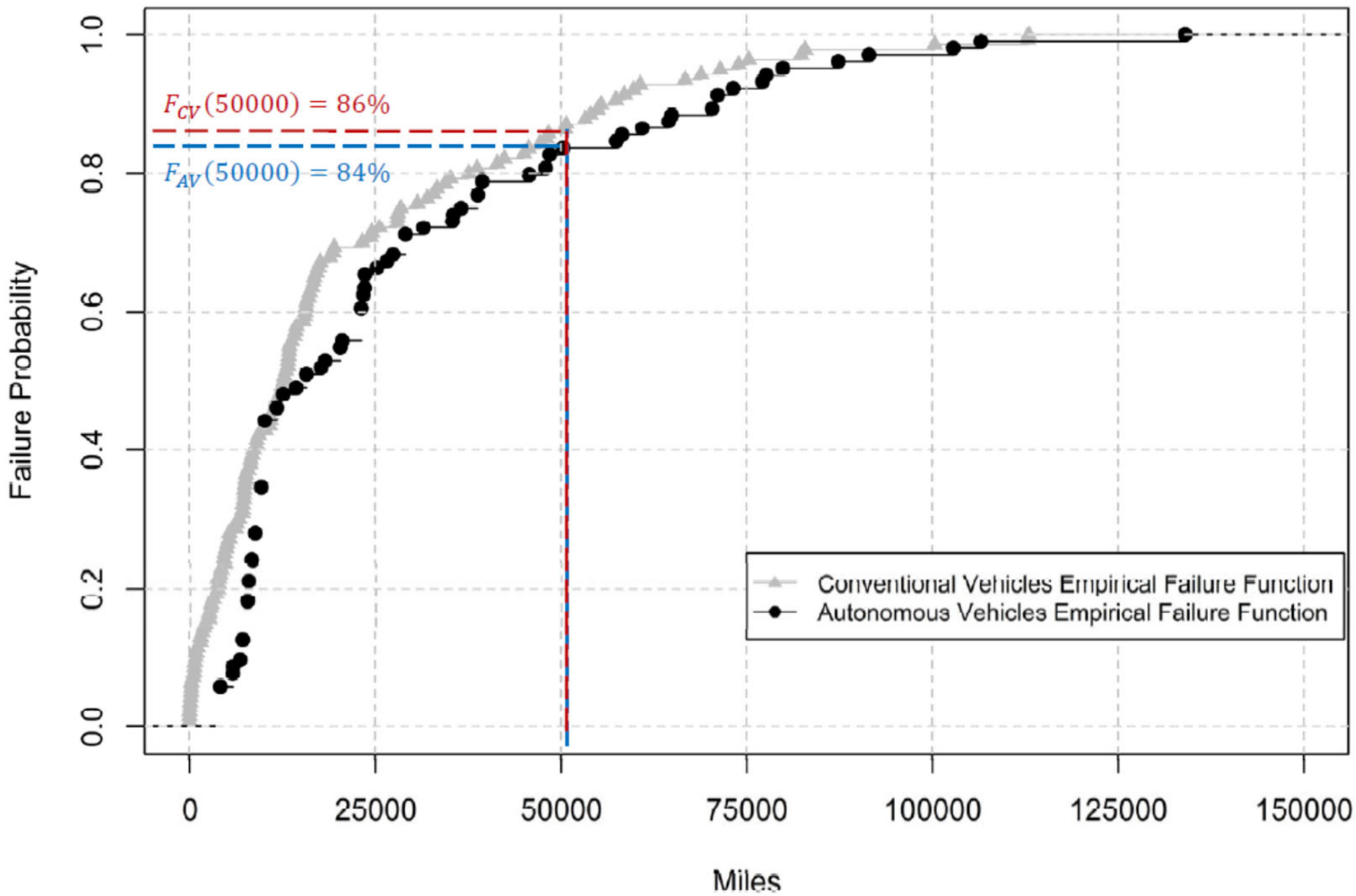
Safety Performance of AVs



Safety Performance of AVs

Mileage	Conventional Vehicle Crash-free Expectancy (miles)	Automated Vehicle Crash-free Expectancy (miles)	Difference in Miles	Difference in Percentage
10,000	3884	7,500	3616	93%
25000	8027	11,724	3697	46%
50,000	13,730	16,863	3133	23%
100,000	20,257	24,813	4556	22%
150,000	21,612	27,399	5787	27%

Safety Performance of AVs



Note: Hazard model is also covered in Chapter 7 of the textbook.