Estimating the Conflict-crash Relationship

Towards Traffic Conflicts Analysis Suitable for Safety Management Practice

Andrew P. Tarko
Professor of Civil Engineering
Director of Center for Road Safety

Purdue University
Lyles School of Civil Engineering
West Lafayette, Indiana, USA

8th Road Safety and Simulation International Conference RSS 2022
Theme: Road Safety and Digitalization
8-10 June 2022, Athens, Greece

Introduction

Two general applications of traffic conflicts:

(1) Quick assessment of crash risk and countermeasures effectiveness (safety engineers)

(2) Safety effects analysis using statistical models (safety researchers)

This presentation provides the ML-based connection of traffic conflicts to crashes and discusses its estimation suitable for safety engineers.

Introduction

Failure-caused traffic events include conflicts and crashes. During a conflict, crash does not occur but is uncomfortably close.

Let *x* be Lomax-distributed *response time* measured from the moment when crash is too close be acceptable to the moment of evasive action.

The corresponding tail (exceedance) distribution of response time is:

$$\bar{F}(x|k,\theta) = 1 - F(x|k,\theta) = (1 + \theta x)^{-k}$$

Let *n* be the number of failure-caused traffic events when time-to-collision is shorter than acceptable x_c.

The expected number of crashes associated with *n* events is:

$$Q_c = n \cdot \bar{F}(x_c) = n \cdot (1 + \theta x_c)^{-k}$$

At crash, response time *x* is truncated at x_c and *x* is not observable in its entirety. Let *c* be the number of crashes.

Introduction

OLS Estimate of k Parameter

Sort n response times x_i in the increasing order.

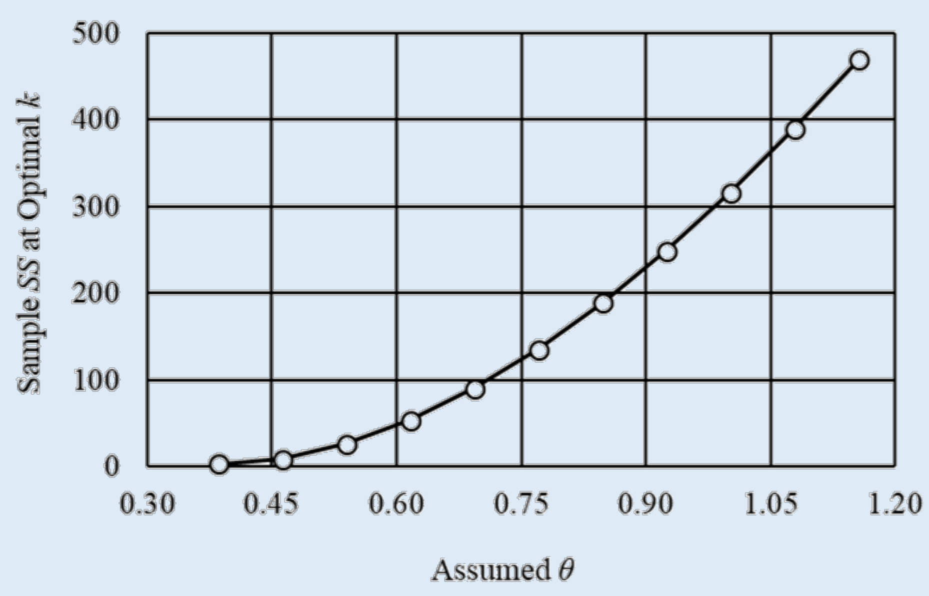
Calculate $\bar{F}(x_i) = 1 - (i - 0.5)/n)$.

Take *ln* of both sides of equation: $1 - (i - 0.5)/n) = (1 + \theta x_i)^{-k}$.

assumed

$$-\ln[1 - (i - 0.5)/n)] = \boxed{k} \ln(1 + \theta x_i) \qquad i = 1 \ldots (n - c)$$

? OLS estimate of k is: $\hat{k} = \dfrac{-\sum_{i=1}^{n-c} \ln(1-(i-0.5)/n)\ln(1+\theta x_i)}{\sum_{i=1}^{n-c}[\ln(1+\theta x_i)]^2}$



Assume $\theta = 1/x_c$?

$$\hat{Q}_c = n \cdot 2^{-\hat{k}}$$

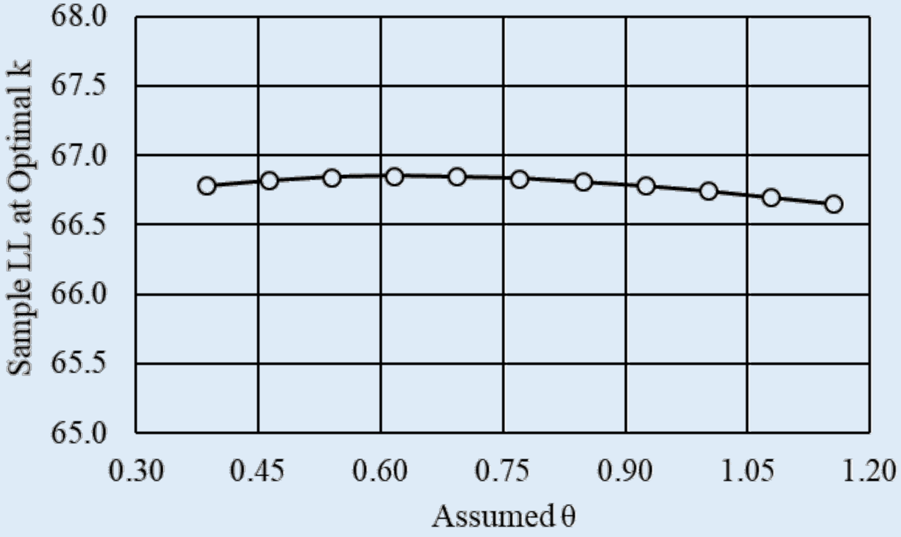
Three Points to Be Addressed

1. ML estimation of k parameter
2. Sensitivity of k and Q_c estimates to assumed θ parameter in single-parameter estimation
3. Two-parameter estimation of Lomax

ML Estimate of k Parameter

LL for c crashes and $(n - c)$ traffic conflicts under x_c threshold is:

$$LL = \underbrace{(n-c)\ln(\theta k) - (k+1)\sum_{i=1}^{n-c}\ln(1+\theta x_i)}_{\text{traffic conflicts}} \underbrace{- ck\ln(1+\theta x_c)}_{\text{crashes}}$$



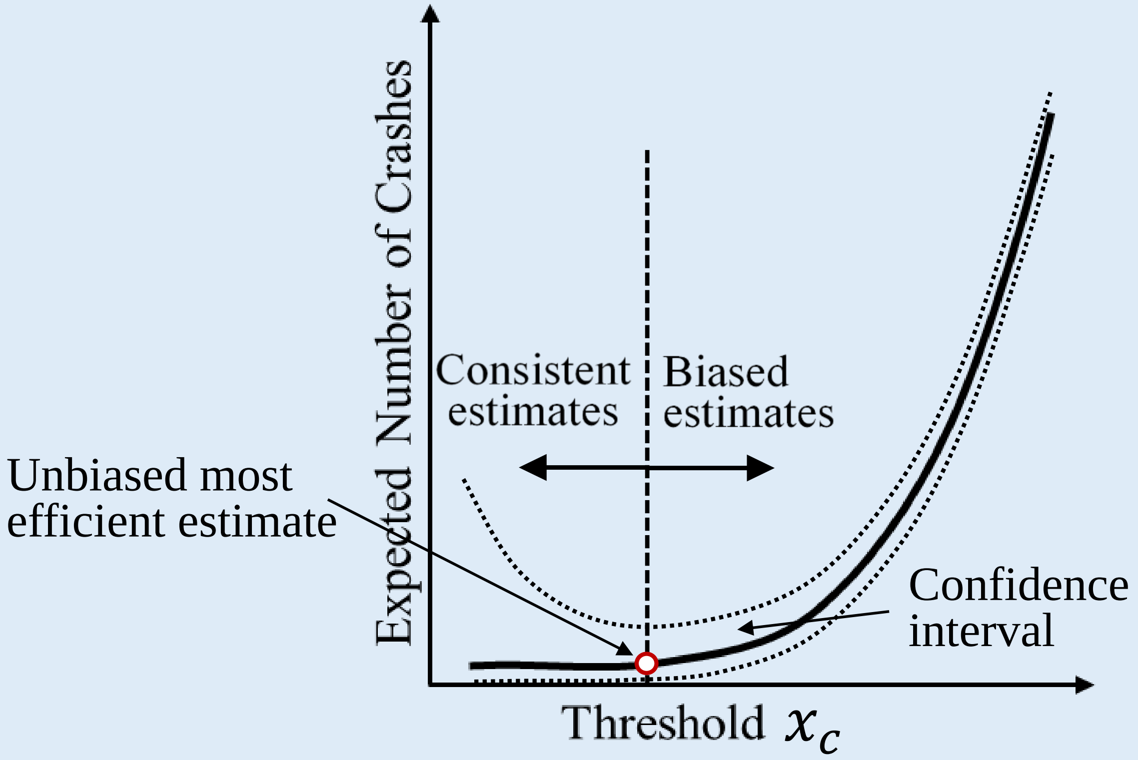
$$\frac{\partial LL}{\partial k} = 0$$ yieds:

$$k = \frac{n-c}{\sum_{i=1}^{n-c}\ln(1+\theta x_i) + c\ln(1+\theta x_c)}$$

For $\theta = 1/x_c$:

$$k = \frac{n-c}{\sum_{i=1}^{n-c}\ln(1+x_i/x_c) + c\ln(2)}$$

Selection of Threshold x_c

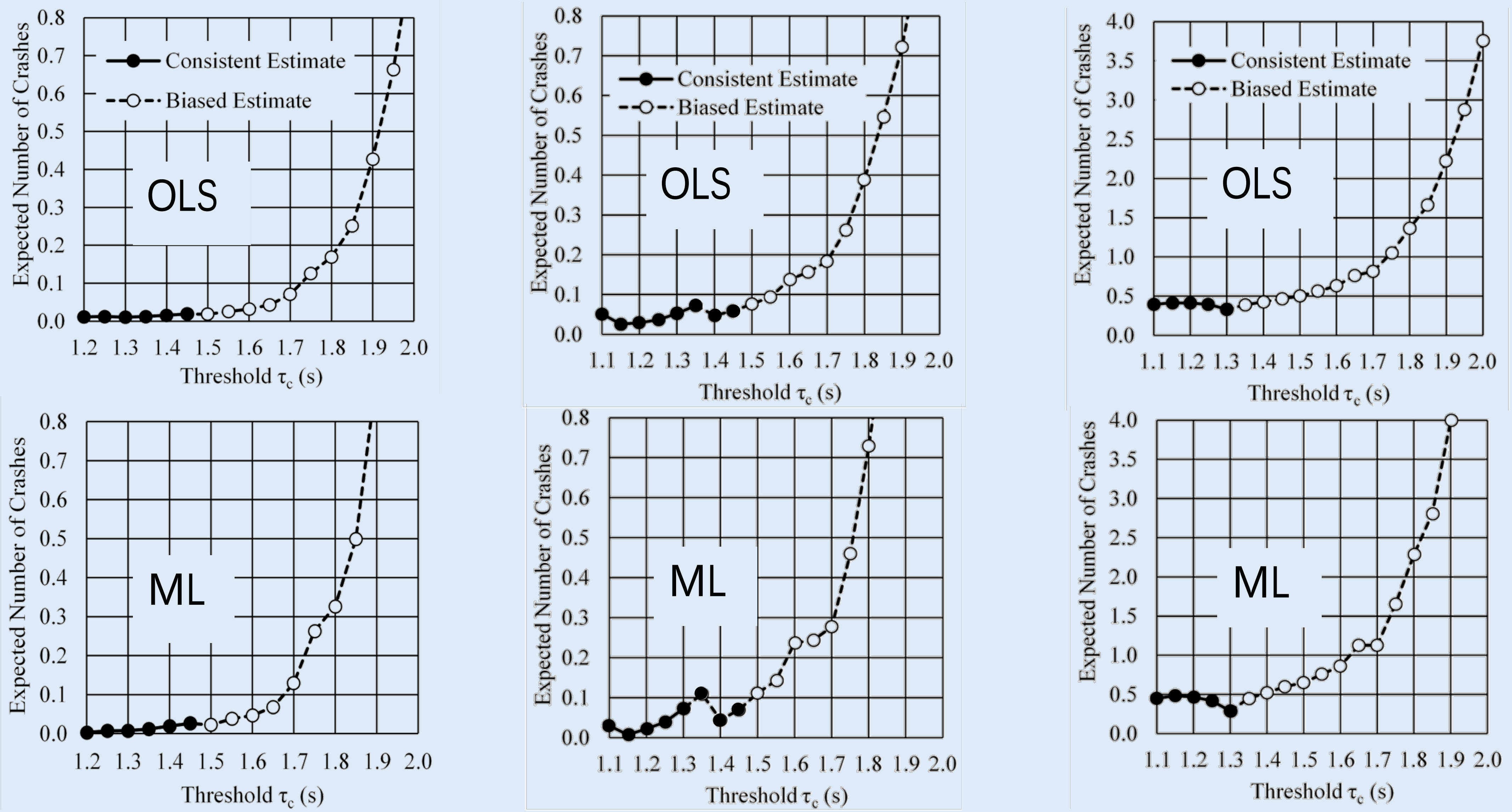


Expected Crashes Profiles

Female drivers 45-64

Male drivers 45-64

Male drivers 16-24



Comparison of OLS and ML Estimates

Rear-end Collisions in of SHRP2

Driver Category	**Method**	**Threshold x_c**	**No. of Conflicts**	**Parameter k**	**Prob. of Crash Given Conflict *PrC***	**Expected Crashes**
Female 45-64	OLS	1.45	47	11.257	0.000408	**0.0192**
	ML	1.45	47	10.797	0.000562	**0.0264**
Male 45-64	OLS	1.45	56	9.885	0.001057	**0.0592**
	ML	1.45	56	9.647	0.001247	**0.0699**
Male 16-24	OLS	1.25	52	7.040	0.007601	**0.3953**
	ML	1.25	52	6.955	0.008059	**0.4191**

Note: The values of threshold x_c were determined with a higher precision than in the original publication (Tarko, 2020).

Comparing ML and OLS Estimates
Simulation Experiments

- The Lomax distributions obtained for the SHRP2 drivers were used in the experiments
- Three sources of variability in crash estimates were studied:
  1) Arbitrary selection of parameter θ (range 0.38 – 1.15)
  2) Randomness (90% conf. interval length)
  3) Driver types (male 16-24, female 45-64)

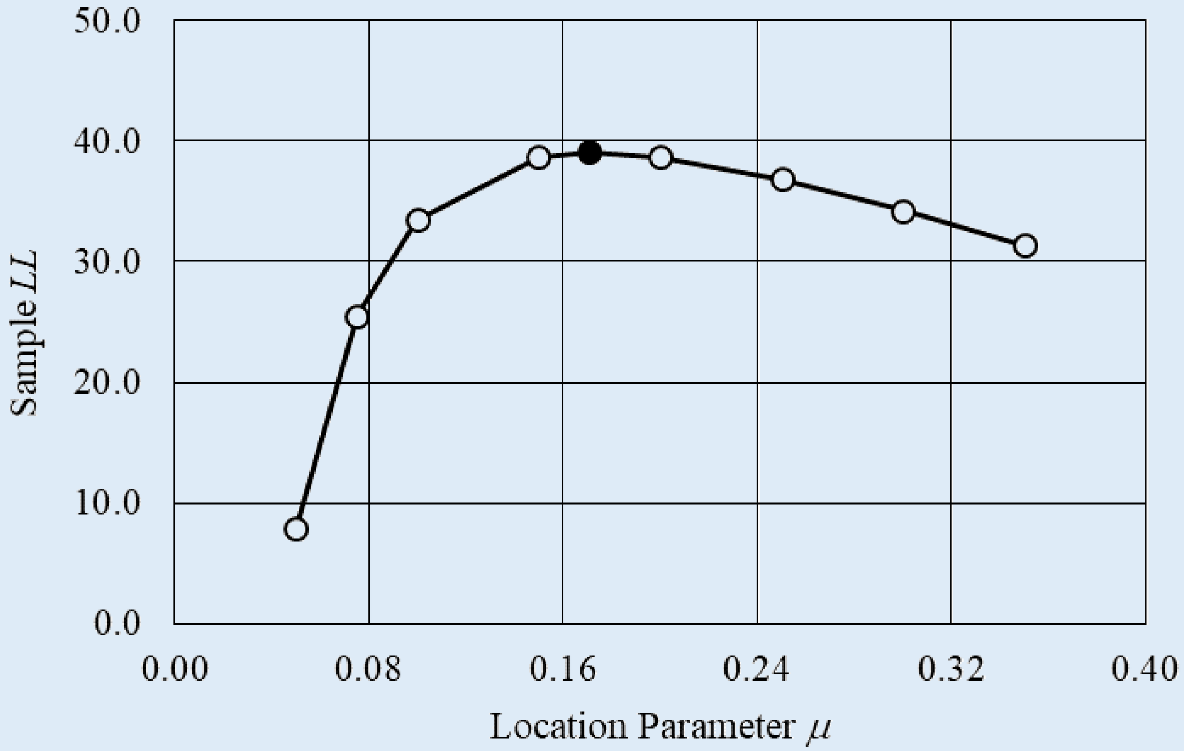
Variability in Expected Number of Crashes Estimates Q_c (spread)

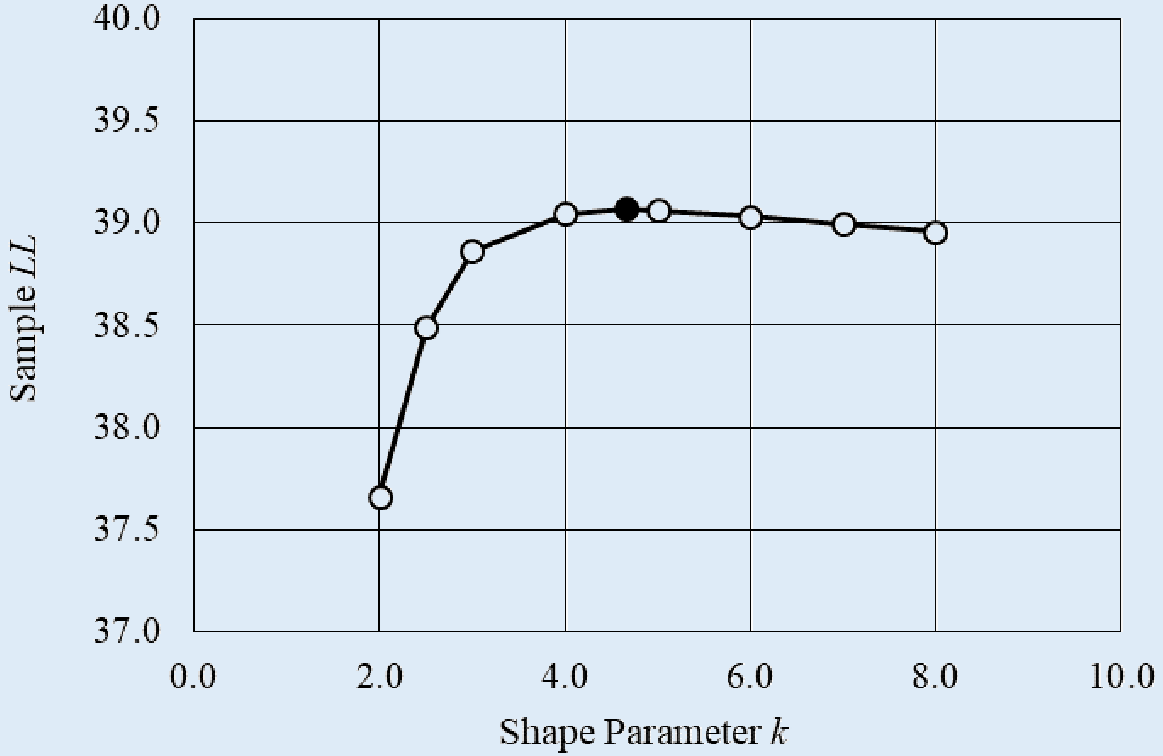
Source of Q_c Estimate Variability	Q_c Spread = $Q_{c,max}$ - $Q_{c,min}$		Notes
	OLS Estimate	ML Estimate	
Parameter θ selected from [0.385, 1.155]	0.673 – 0.390 = **0.283**	0.709 – 0.298 = **0.411**	Average from 200 trials to control randomness
Randomness (90% confidence interval)	1.024 – 0.163 = **0.861**	1.007 – 0.153 = **0.854**	Sample of 100 conflicts
Driver type (male 16-24 vs. female 45-64)	0.527-0.067 = **0.460**	0.494 – 0.062 = **0.432**	Fixed exposure (50,000 miles of car following)

Reparametrized Lomax Distribution

Altun (2021)

$$\theta = \frac{1}{\mu(k-1)} \qquad f(x) = \frac{k}{\mu(k-1)}\left(1 + \frac{x}{\mu(k-1)}\right)^{-k-1}$$





Concluding Remarks

Neither ML nor OLS methods are efficient in estimating both the Lomax parameters. This finding concurs with the published research.

The existing OLS and ML single-parameter methods with θ set at x_c seems to be acceptable for estimating the crash frequency.

The reparametrized Lomax distribution seems to be promising but estimates of the shape parameter k exhibit strong variance. More research is needed.

Currently, estimation efficiency as more important than estimation consistency. Massive data potentially available in the future may increase the importance of the estimation consistency.

This presentation did not discuss accounting for heterogeneity via regression analysis to avoid its estimation complexity.

Thank you