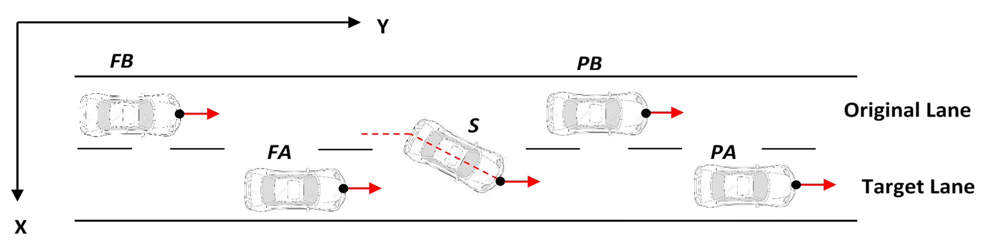
Comparisons of Mandatory and Discretionary Lane Changing Behavior on Freeways

Matthew Vechione, Esmaeil Balal, and Ruey Long Cheu
The University of Texas at El Paso

Background

- A lane change is a lateral movement of a vehicle which is always accompanied with a longitudinal movement.
- A lane changing event involves up to five vehicles (see *S*, *FB*, *PB*, *FA*, *PA* in the figure below).



- A lane change may be modeled as a four-step process:
  (1) motivation;
  (2) selection of target lane;
  (3) **checking for opportunity to move**; and
  (4) the actual move.
  This research focuses on step (3).
- There are two types of lane changes on freeways: mandatory and discretionary.
  - A Mandatory Lane Change (MLC) occurs when a driver **must** change lanes to exit a freeway, avoid a lane closure downstream, turn at a downstream intersection, etc.
  - A Discretionary Lane Change (DLC) occurs at a driver's own **discretion** for faster speed, greater following distance, further line-of-sight, etc.
- A driver is expected to have different decision rules and/or risk-taking behavior for the two types of lane changes.

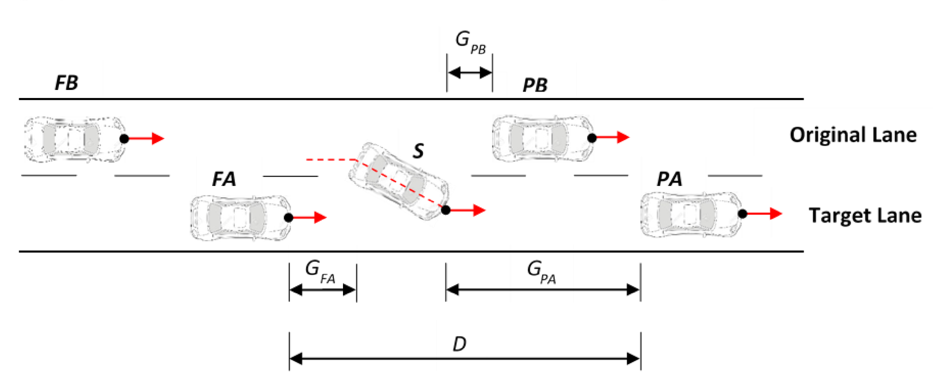
Objectives

The objectives of this research are to:

1. Examine descriptive statistics for variables that describe vehicle interactions for MLCs and DLCs, respectively.
2. For each variable, conduct hypothesis test on the difference between the means of MLCs and DLCs.
3. For each variable, apply the Kolmogorov-Smirnov (KS) test to test the difference in the *observed* cumulative probability distributions between MLCs and DLCs.
4. For each variable, fit the probability distributions to the MLC and DLC data respectively, and use the KS test to test the difference between the *fitted* probability distributions.

Literature Review

- Based on a survey from 443 drivers in El Paso, TX by *Balal et al. (2014)*, the top four input parameters were gaps and distances, as shown in the figure below.



Notation	Definition	Unit	Range
G_{PB}	Gap between *S* and *PB*	m	≥0
G_{PA}	Gap between *S* and *PA*	m	≥0
G_{FA}	Gap between *S* and *FA*	m	≥0
D	Gap between *PA* and *FA*	m	≥0

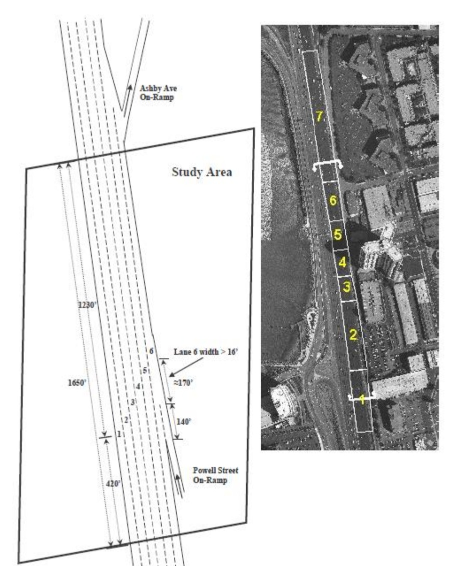
- Formulas for calculating the lane changing variables are:
  - Front gap before lane change (in meters):
    $G_{PB} = (Y_{PB} - L_{PB}) - (Y_S), \;\; G_{PB} \geq 0$
  - Front gap after lane change (in meters):
    $G_{PA} = (Y_{PA} - L_{PA}) - (Y_S), \;\; G_{PA} \geq 0$
  - Rear gap after lane change (in meters):
    $G_{FA} = (Y_S - L_S) - (Y_{FA}), \;\; G_{FA} \geq 0$
  - Distance (in meters):
    $D = (Y_{PA} - L_{PA}) - (Y_{FA}), \;\; D \geq 0$

  where:
  - *L* is the length of the vehicle;
  - *Y* is the longitudinal position of each vehicle;
  - *P* represents a preceding vehicle;
  - *F* represents a following vehicle;
  - *B* is before the lane change; and
  - *A* is after the lane change.

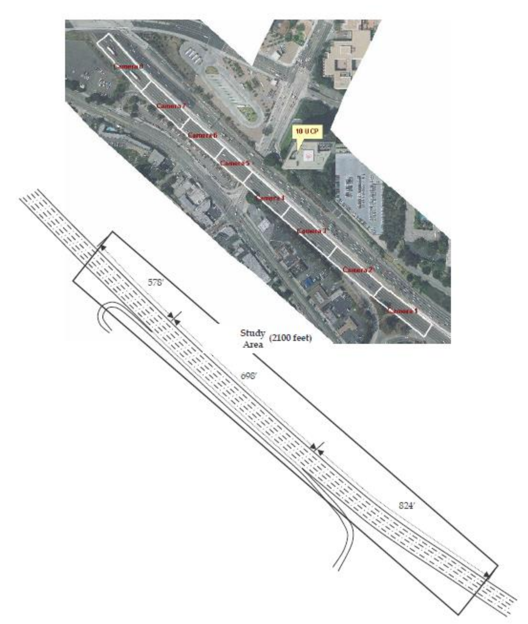
Vehicle Trajectory Data

- From NGSIM data base.

Interstate 80 (I-80)
Emeryville, CA
Dataset A



U.S. Highway 101
Los Angeles, CA
Dataset B



Cambridge Systematics, Inc. (2005)

Methodology

- Only passenger cars selected as subject vehicles;
- Vehicles that changed lanes between lanes 5 and 6 were assumed to make a MLC;
- Vehicles that changed lanes between lanes 2 to lane 4 were assumed to make a DLC;
- Lane 1 omitted, as it is a HOV lane;
- For each subject vehicle, the time *t* when the lane changing event occurred was taken as time when the front center of the subject vehicle crossed the lane markers;
- Variable values were calculated at *t*-0.4, *t*-0.3, *t*-0.2, *t*-0.1, and *t* seconds, and the average values from *t*-0.4 to *t* seconds were used as the representative value. The averaging of data to 0.5 second intervals was to:
  i. Reduce error caused by instantaneous values in the NGSIM data;
  ii. Be more consistent with human perception time; and
  iii. Be consistent with other research that used NGSIM data.

Statistical Analyses

1. Descriptive Statistics

- **Dataset A**

Variable	G_{PB}		G_{PA}		G_{FA}		*D*	
Unit	m		m		m		m	
	MLC	**DLC**	**MLC**	**DLC**	**MLC**	**DLC**	**MLC**	**DLC**
Sample size	166	135	166	135	166	135	166	135
Min	0.61	4.33	0.31	0.07	1.40	0.49	10.25	6.06
Max	124.26	76.97	47.75	105.37	80.03	93.07	115.33	162.15
Mean	15.08	15.18	10.32	11.46	15.35	17.58	30.12	33.42
Std. deviation	13.97	8.63	8.66	13.43	12.10	14.66	16.65	20.85
Skewness	4.07	3.13	1.97	3.44	2.11	1.94	1.91	2.65

- **Dataset B**

Variable	G_{PB}		G_{PA}		G_{FA}		*D*	
Unit	m		m		m		m	
	MLC	**DLC**	**MLC**	**DLC**	**MLC**	**DLC**	**MLC**	**DLC**
Sample size	71	128	71	128	71	128	71	128
Min	5.63	3.79	3.46	0.82	1.93	0.45	11.51	15.24
Max	185.81	74.08	160.89	216.07	92.06	103.51	172.54	234.74
Mean	50.74	19.11	22.44	20.83	22.88	20.95	49.72	46.06
Std. deviation	40.70	12.72	24.01	24.82	18.69	16.10	30.80	27.49
Skewness	1.11	2.06	3.35	4.65	1.59	2.37	1.64	3.34

2. Difference Between Two Means

- **Dataset A**

Variable	G_{PB}		G_{PA}		G_{FA}		*D*	
Unit	m		m		m		m	
	MLC	**DLC**	**MLC**	**DLC**	**MLC**	**DLC**	**MLC**	**DLC**
Sample size	166	135	166	135	166	135	166	135
Mean	15.08	15.18	10.32	11.46	15.35	17.58	30.12	33.42
Std. deviation	13.97	8.63	8.66	13.43	12.10	14.66	16.65	20.85
t-statistic	-0.07		-0.85		-1.42		-1.49	
p-value	0.944		0.398		0.157		0.137	
Conclusion (α/2=0.025)	Fail to reject H_0		Fail to reject H_0		Fail to reject H_0		Fail to reject H_0	

- **Dataset B**

Variable	G_{PB}		G_{PA}		G_{FA}		*D*	
Unit	m		m		m		m	
	MLC	**DLC**	**MLC**	**DLC**	**MLC**	**DLC**	**MLC**	**DLC**
Sample size	71	128	71	128	71	128	71	128
Mean	50.74	19.11	22.44	20.83	22.88	20.95	49.72	46.06
Std. deviation	40.70	12.72	24.01	24.82	18.69	16.10	30.80	27.49
t-statistic	6.38		0.45		0.73		0.83	
p-value	0.000		0.654		0.466		0.406	
Conclusion (α/2=0.025)	**Reject H_0**		Fail to reject H_0		Fail to reject H_0		Fail to reject H_0	

3. Observed Distributions

- The KS test compares a *cumulative distribution* against the theoretical cumulative distribution or two *cumulative distributions* against one another.
- The maximum difference between the two distributions is computed by:

$$d = \max_{x,y} |F(x) - F(y)|$$

- *d* is compared to a critical value $d_{n_1,n_2,\alpha}$:

$$d_{n_1,n_2,\alpha} = k_\alpha \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

  where n_1 and n_2 are the sample sizes of the two distributions, k_α is the KS test parameter with level of significance = α.

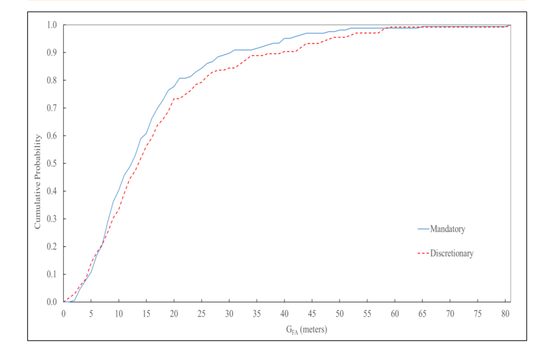
- **Dataset A**

Variable	G_{PB}	G_{PA}	G_{FA}	*D*
d	0.164	0.126	0.076	0.085
α=0.10 Critical Value = 0.142	**Reject H_0**	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0
α=0.05 Critical Value = 0.157	**Reject H_0**	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0

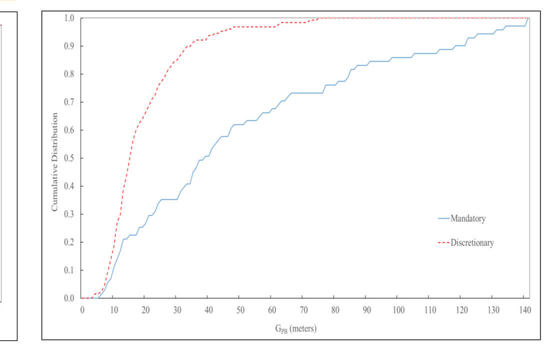
- **Dataset B**

Variable	G_{PB}	G_{PA}	G_{FA}	*D*
d	0.499	0.101	0.108	0.146
α=0.10 Critical Value = 0.181	**Reject H_0**	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0
α=0.05 Critical Value = 0.201	**Reject H_0**	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0

Best case (G_{FA} in Dataset A)



Worst case (G_{PB} in Dataset B)



4. Fitted Probability Distributions

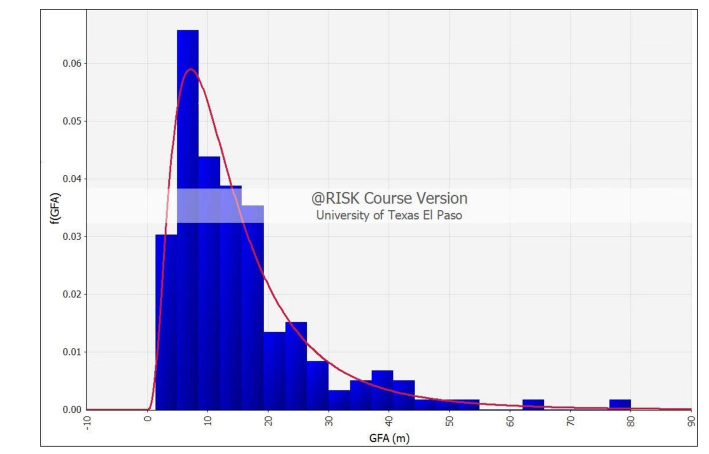
- The observed data was then fitted with probability distributions using @RISK.
- MLC

Variable	G_{PB}	G_{PA}	G_{FA}	*D*
Dataset A				
Unit	m	m	m	m
Best fit	Log-logistic	Pearson 5	**Log-normal**	Gamma
2nd best fit	Pearson 5	**Log-normal**	Inverse Gaussian	Inverse Gaussian
3rd best fit	**Log-normal**	Inverse Gaussian	Pearson 5	**Log-normal**
Recommended	Log-normal			
Log-normal location parameter, λ	2.404	2.068	2.489	3.272
Log-normal scale parameter, ξ	0.787	0.730	0.695	0.516
Dataset B				
Unit	m	m	m	m
Best fit	Exponential	Inverse Gaussian	Inverse Gaussian	Inverse Gaussian
2nd best fit	Inverse Gaussian	**Log-normal**	**Log-normal**	Gamma
3rd best fit	**Log-normal**	Pearson 5	Pearson 5	**Log-normal**
Recommended	Log-normal			
Log-normal location parameter, λ	3.678	2.729	2.874	3.744
Log-normal scale parameter, ξ	0.705	0.873	0.715	0.570

- DLC:

Variable	G_{PB}	G_{PA}	G_{FA}	*D*
Dataset A				
Unit	m	m	m	m
Best fit	Pearson 5	Exponential	Inverse Gaussian	Log-logistic
2nd best fit	Log-logistic	**Log-normal**	**Log-normal**	Pearson 5
3rd best fit	**Log-normal**	Inverse Gaussian	Pearson 5	**Log-normal**
Recommended	Log-normal			
Log-normal location parameter, λ	2.580	2.006	2.603	3.345
Log-normal scale parameter, ξ	0.529	0.930	0.726	0.573
Dataset B				
Unit	m	m	m	m
Best fit	Pearson 5	**Log-normal**	Log-logistic	Log-logistic
2nd best fit	**Log-normal**	Log-logistic	Pearson 5	Pearson 5
3rd best fit	Log-logistic	Pearson 5	**Log-normal**	**Log-normal**
Recommended	Log-normal			
Log-normal location parameter, λ	2.767	2.594	2.810	3.678
Log-normal scale parameter, ξ	0.606	0.940	0.681	0.552

- The lognormal distribution was recommended.
- Example: log-normal distribution fitted for G_{FA}, Dataset A:



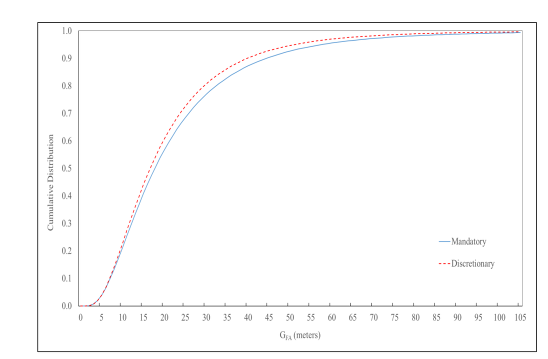
- The KS test was then applied to the fitted log-normal distributions for MLC and DLC.
- **Dataset A** – difference between MLC and DLC

Variable		G_{PB}	G_{PA}	G_{FA}	*D*
d		0.168	0.077	0.066	0.066
α= 0.10	Critical Value	0.155	0.165	0.178	0.135
	Conclusion	**Reject H_0**	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0
α= 0.05	Critical Value	0.172	0.183	0.197	0.150
	Conclusion	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0

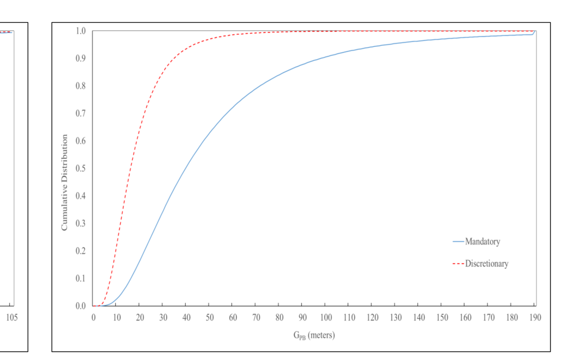
- **Dataset B** – difference between MLC and DLC

Variable		G_{PB}	G_{PA}	G_{FA}	*D*
d		0.515	0.066	0.041	0.049
α= 0.10	Critical Value	0.126	0.117	0.169	0.113
	Conclusion	**Reject H_0**	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0
α= 0.05	Critical Value	0.139	0.129	0.187	0.125
	Conclusion	**Reject H_0**	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0

Best case (G_{FA} in Dataset B)



Worst case (G_{PB} in Dataset B)



Conclusions

- All variables may be described by the log-normal distribution.
- There is no significant difference between MLC and DLC for the three variables in the **target lane** (i.e. G_{PA}, G_{FA}, and *D*).
  - These may be common variables between MLCs and DLCs
- For G_{PB} (in the original lane), significant differences are found between MLC and DLC:
  - Population means between MLCs and DLCs in Dataset B (at 95% confidence)
  - *Observed* probability distributions in Dataset A (at 95% confidence)
  - *Observed* probability distributions in Dataset B (at 95% confidence)
  - *Fitted* log-normal distributions in Dataset A (at 90% confidence)
  - *Fitted* log-normal distributions in Dataset B (at 95% confidence)