

Investigation the Stochastic behaviour of the Traffic Flow: A Case Study of a Section of a Road

Mehboob Ali Jatoi^{1*}, Shakeel Ahmed Kamboh², Oshaque Ali Abro³, Saeed Ahmed Rajput^{2,4}, Liaquat Ali Zardari⁵

^{1*}Department of Basic Science and Related Studies, Quaid-e-Awam University of Engineering, Science and Technology Nawabshah Sindh Pakistan; ²Department of Mathematics and Statistics, Quaid-e-Awam University of Engineering, Science and Technology Nawabshah Sindh Pakistan; ³Dawood university of Engineering and Technology, Karachi; ⁴Department of Basic Science and Related Studies, The University of Larkano; ⁵Faculty of Haritage, Aror University of Art, Architecture, Design and Haritage Sukkur

Keywords: Traffic flow, Goodness-of-fit, Probability Distributions, Nawabshah. **Subject Classification:** Statistics, Distributions.

Journal Info:

Submitted:
March 27, 2024
Accepted:
April 12, 2024
Published:
May 04, 2024

Abstract The stochastic behavior is a key factor influencing the current state of vehicle flow in real-time traffic. This paper investigates the stochastic behavior of real-time traffic flow for a section of road using probability distribution fitting. Traffic flow data is collected through manual counts conducted at various inflow and outflow points within the targeted area. One specific path within this area, characterized by significant traffic congestion, is chosen for data collection due to its potential to reveal stochastic parameters. Data collection occurs over a week, from 7:00 to 19:00, at Nawabshah, Pakistan. MATLAB is utilized for fitting different distributions such as Normal, Lognormal, Weibull, Gamma, and Exponential, using probability plots of traffic flow data. Goodness-of-fit tests, including Kolmogorov-Smirnov, modified Kolmogorov-Smirnov, and Anderson-Darling tests, are conducted at a 95% confidence level to determine acceptance or rejection of hypotheses. The hypotheses are accepted for Normal, Weibull, and Gamma distributions based on the test results. Among these three accepted distributions, the lognormal probability distribution demonstrates the best fit, as indicated by the rank of p-values from the hypothesis tests.

***Correspondence Author Email Address:**

mehboobali.quest@gmail.com

1 INTRODUCTION

Traffic flow of vehicles over the road is depends on the congestion or free flow, in developed countries the intelligent transportation system is work for the betterment of road condition for the traffic flow. The dynamical system of the vehicles changed with changed with congestion or non congestion on section of road. The nonlinear behavior of the vehicles occur mainly with congestion and in nonlinear condition of the vehicles the shock waves creates on the position of the congestion, so the behavior of the traffic may be concern as stochastic[1, 3, 4].

Analysis of traffic data strongly depends on its distribution pattern, using the traffic data a best fit probability distribution can be used to find a nature of traffic flow over the section of road. The data is collected manually for a week from 7:00 to 19:00 to count the vehicles over the location Nawabshah Pakistan [2].

The distribution of fit using goodness-of-fit will take place for real data of traffic flow the different tests of hypothesis will be applied using MATLAB distribution fit and different hypothesis tests like Kolmogorov-Smirnov, Kolmogorov-Smirnov modified and Anderson-Darling by considering the p-values of test for hypothesis acceptance or rejections [7, 8, 10].

The traffic flow condition is depends over the distribution of traffic speed of the vehicles and speed of vehicle is depend over the situation of traffic condition of the section of road, the speed will be maximum over the no congestion of traffic and minimum with congestion traffic. Maqhrour[6, 9, 32] used the traffic flow for budapest (Hungary) for the best fit over the collected data and found the normal, exponential, lognormal, gamma and chi-Square are fit using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). DS Berry (1951) used the data and speed for the best relation between the data and speed distribution over the section of road, [11, 13, 19] studied the superstatistical analysis of traffic flow and found the beta distribution is the best fit with a small fluctuation with chi-square test for the traffic flow data, [18, 26] used the traffic data with condition the data is normally distribution and used to check the data is unimodal or a bimodal and found the traffic data follow the unimodal only, [22, 30] used to estimate the travel time distribution using all the distribution to estimate the time distribution under the congestion and free flow congestion, [1, 24] used the data for superstatistical analysis with best fit data of speed vehicles flow, [23] used the real traffic data and estimate the statistical distributions for forecasting under the uncertainty condition of traffic condition, [5] used the studied the statistical analysis of traffic flow using lognormal distribution of traffic flow,[12] use the time-gap traffic data under the mix-traffic condition and used all the distribution by combining two distribution and also used the goodness-of-fit distribution with hypothesis test, [27] use the time headway and speed headway under the mix condition of traffic flow and test for the statistical analysis using distribution fit,[15, 16] studied for the determination of best fit probability distribution of rainfall data in Bangladesh, and test the data under different hypothesis test and distribution technique, [26] use the regression analysis for the fit the probability distribution over the traffic data by considering the traffic data into different types at the location of Hong Kong, [30] used the distribution fit by probability distribution function over the Origin-Destination of traffic network for the congested traffic flow over the considered network and used the Generalized method of Moment with exact and approximate estimator [20] used the goodness-of-fit probability distribution for the traffic data for analysis the traffic network tail and used different hypothesis tests for goodness-of-fit data and Monte Carlo simulation, [14] statistical analyze the traffic data collected over the traffic network for the modeling and simulation purpose and for the normality of the traffic network data.

[21] discussed the methods for data collection, Pneumatic road tubes, sensors, manually count, survey of road, radar system, video recording and detection, GPS based traffic detection, smart phone apps as GPS.

[17] used the randomly collected data using sensor which is high costly and internet based, [17, 31] used the tracking the pedestrian on the road by neglecting the traffic vehicles using the video recording method. [23, 29] has analyze into the traffic data collected method and described the manual count method and described the manual method is low casting and risk-less technique to collect the data due to collection of specific vehicles, whiles in other method of collecting data, all types of vehicles including human movement is also count as traffic data. In this research the method of data collecting is manual method to collect the data on one path for a week. Most of the researchers has done the statistical data analysis and distribution fit for the analysis and types of traffic flow over the traffic data, this study capture the traffic flow data of Nawabshah Pakistan, and then used it to find the best fit probability distribution for the type of traffic flow.

2 METHODOLOGY

2.1 AREA SELECTION

Figure 1 shows Satellite view of the section of road of Shaheed Benazir Abad Nawabshah Sindh Pakistan. The road is with the 2.0 miles starting from New Naka to Zero Point Nawabshah, the starting point is consist of 8 lanes we consider data for one lane, the vehicles may move to the considered lane from any lane, the road is in with ramps on and road off, but we consider the road with no ramps on and off because the road data should be collected at the fixed location so the flow from on ramp and off ramp may affect at the targeted location. So the road is smooth with no ramp in or ramp out so the inflow and out flow of the lane will be same. The traffic flow data is collected manually by visiting the located place using vehicles count, the vehicles consist of all type of vehicles including bikes, cars and heavy vehicles. The data is collected into the one hour of interval, like the total number of vehicles in one hour, and same is collected for free flow time interval and peak hours.



Figure 1. Satellite view of section of road of Nawabshah, Shaheed Benazir Abad

2.2 DATA-COLLECTION

For real data collection over the section of road a manual method is adopted by visiting the actual place of road, the data is collected for one day continuously from 7:00 to 19:00 for a complete week of the month from January 15, 2024 to January 21, 2024 because it including the effect of peak hours, peak days at the main inflow and outflow of the section of road. The vehicles are to be considered as volume of traffic including all types of traffic as car, bikes, cycles and heavy vehicles because there is no any restriction for the enters or exists of any types of vehicles because the section of road is used for all types of vehicles. Table 1 shows the total number of vehicles flow over the section of road.

Table 1 Vehicles flow data collection over the section of road.

	Collection of data over the section of road with no ramp on or ramp off												Date: January 15-19, 2024					Statistics Hour-wise				
	Day/Time	7:00-8:00	8:00-9:00	9:00-10:00	10:00-11:00	11:00-12:00	12:00-13:00	13:00-14:00	14:00-15:00	15:00-16:00	16:00-17:00	17:00-18:00	18:00-19:00	Total	Average	Standard Devi	Range	Skewness				
Monday	198	313	412	431	466	501	477	422	391	242	297	202	4352	362.67	108.12	303.00	-0.42					
Tuesday	242	323	348	411	512	628	492	455	402	376	310	232	4731	394.25	115.30	396.00	0.48					
Wednesday	233	297	334	377	389	428	511	459	421	368	318	294	4429	369.08	78.65	278.00	0.10					
Thursday	311	432	512	623	602	584	487	392	429	512	411	334	5629	469.08	101.85	312.00	0.05					
Friday	263	366	412	427	467	535	512	499	474	428	329	276	4988	415.67	90.25	272.00	-0.53					
Saturday	248	345	411	518	511	476	433	487	389	388	413	377	4996	416.33	76.69	270.00	-0.67					
Sunday	152	373	375	287	390	344	516	412	367	365	310	311	4202	350.17	85.81	364.00	-0.56					
Total	1,647.00	2,449.00	2,804.00	3,074.00	3,337.00	3,496.00	3,428.00	3,126.00	2,873.00	2,679.00	2,388.00	2,026.00	33,327.00	396.75	14.89	126.00	0.67					
Statistics, day-wise	Average	235.29	349.86	400.57	439.14	476.71	499.43	489.71	446.57	410.43	382.71	341.14	289.43	79.51	238.13							
	Standard	50.07	45.54	58.65	106.49	74.77	95.53	28.96	39.55	34.86	80.80	49.35	59.53	41.20		28.04						
	Range	159.00	135.00	178.00	336.00	213.00	284.00	83.00	107.00	107.00	270.00	116.00	175.00	391.00			258.50					
	Skewness	-0.32	0.87	1.08	0.56	0.40	-0.36	-1.40	-0.02	0.91	-0.24	1.07	-0.10	1.22				0.94				

Table 1 shows the total number of vehicles for a week from Monday to Sunday, from 7:00 to 19:00 with the interval of one hour. The data is statistically tested with total number of vehicles days-wise and week-wise.

The average flow of vehicles is calculated according to day and week-wise, in figure 3, the red color data shows the total week-wise average, standard deviation, range and skewness, blue data shows the week-wise per hour average, standard deviation, range and skewness, while the diagonal values shows the total average, standard deviation, range and skewness of the collected data. In order to use a suitable distribution over the collected data for a best fit, the Normal distribution, Lognormal distribution, Exponential distribution, Gamma Distribution, Weibull distribution will be used using MATLAB distribution fit by plot the data in probability-plot (pp-plot) by probability distribution over the collected data. To analyze the stochastic behavior of the traffic flow over the collected data, the probability distribution are suitable form, the behavior of the probability distribution over the collected data is to be consider as the stochastic behavior of the traffic flow over the collected data. The probability distribution function for the defined distributions found in [15].

Table 2 Probability Density Function of different Distributions

Sr	Distribution	Probability distribution function	Range
1	Normal	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \times \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$-\infty \leq x \leq \infty$
2	Lognormal	$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \times \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$	$0 \leq x \leq \infty, -\infty \leq \mu \leq \infty, \sigma > 0$
3	Weibull	$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right)$	$x \geq 0$
4	Gamma	$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)$	$0 \leq x \leq \infty, \alpha \geq 0, \beta \geq 0$
5	Exponential	$f(x) = \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right)$	$0 \leq x \leq \infty, \mu \neq 0$

2.3 GOODNESS-OF-FIT TESTS

The goodness-of-fit is the key to discuss the statistical test to check the behavior of data is fit to the specific distribution or not. It is a type of test which can be applied on the described confidence of interval, and gives the test that the data is accepted for the defined confidence interval or not by two values H_0 and H_1 . The main purpose of this study is to fit a best distribution over the traffic flow data, the data was collected over the section of road of Nawabshah Pakistan. The goodness-of-fit is ever for the hypothesis test accept or rejected H_0 is used when a hypothesis test accept the test and H_1 used when a hypothesis rejected. The three goodness-of-fit tests like Kolmogorov-Smirnov test, Kolmogorov-Smirnov modified test and Anderson-Darling will be used for the distribution with 95% of confidence level. The tests describe the statistics value, p-value for the distribution and description for hypothesis accepts or can't accept the hypothesis will be discussed. The following three tests are to be considered for the goodness-of-fit and the test which accepts the all types of hypothesis test is to be considered as the most fitted for the distribution.

2.4 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test [8] described based over the Empirical distribution function, used to check the data is follow the specific distribution or not and is define as,

$$E_N = \frac{n(i)}{N} \quad (1)$$

Where $n(i)$ is the number of points less than Y_i and Y_i is the points from smallest to largest data point, the statistical test of Kolmogorov-Smirnov test can be defined as,

$$KS(\mu, \sigma) = D = \text{Max}_{1 \leq i \leq N} \left(F(Y_i) - \frac{i-1}{N}, \frac{1}{N} - F(Y_i) \right) \quad (2)$$

Where μ is the mean and sigma is the standard deviation. F is the cumulative distribution function, D is the statistical value and the hypothesis test reject the distribution if the D value greater than critical value comes from table of K-S test.

2.5 Kolmogorov-Smirnov Modified Test

The Kolmogorov-Smirnov Modified test can be defined using [25], used to find the distribution fit for the collected data when it is not normally distributed. The K-S modified test is very much challengeable when the data is much near to normally distribution because the critical values are most probably same values for the K-S test and K-S modified test. The Modified K-S test using the equation

$$KS(\bar{\mu}, \bar{\sigma}) = D = \text{Max}_{1 \leq i \leq N} \left([1 - F(Y_i)] \left[1 - \frac{i-1}{N}\right], \left[1 - \frac{1}{N}\right] - [1 - F(Y_i)] \right) \quad (3)$$

2.6 Anderson -Darling Test

The Anderson-Darling test [28] The Anderson-Darling AD test is used when the collected data follow the specific distribution, the main difference between the K-S test and AD test that K-S test does not depend upon the distribution or says it is not necessary that the data follow the distribution it may be follow or not, while the AD test guarantee about the data must be comes from the specific distribution because the AD test follow the critical value of the hypothesis test. But this test needs much sensitive while using AD test and is used each time different critical value of distribution while using the specific distribution, like by using normal distribution the AD test calculate the critical value while using other distribution same critical value will not be applied on the others distributions, the critical values of the normal distribution, lognormal distribution, Weibull distribution, gamma distribution and exponential distribution are in used for the AD test. For this test the test is one sided it means that distribution follow the one sided tail shape. The AD test is defined using the formula.

$$A^2 = -N - \frac{1}{N} \sum_{i=1}^N (2i-1) [\ln(f(x_i)) + \ln(1 - F(x_{N-i+1}))] \quad (4)$$

Where F is the cumulative distribution function for the distribution. N is the total number of data points. A is the tested value for the AD test, and the distribution is fail to fit over the data if the A value is greater than the critical value of AD test.

2.7 Chi-Square Test

The Chi-Square test is one of the statistic test for the distribution and used when the Collected data is large, then Chi-Square is good for the approximate, this test is mostly used while predicting the collected data. The Chi-Square can be defined using the following equation.

$$\chi^2 = \sum_{j=1}^k \left(\frac{O_j - E_j}{E_j} \right) \quad (5)$$

Where O_j is the observed frequency and E_j is the expected frequency and $E_j = F(x_2) - F(x_1)$ where F is cumulative distribution CDF function of the specific distribution which tested for the fit, k is the total number of the observed value and can be defined as $k = 1 + \log_2^n$, where n is the total length of sample. Here the Chi-Square test is not applied because this study is used to check the stochastic behavior of traffic flow by fitting the suitable distribution over the collected data, while the Chi-Square test is mostly used for predicting the collected data.

3 RESULTS AND DISCUSSION

The figure 1 shows the probability plot of the collected data, the data is considered the probability distribution and converted into the boundary 0 and 1 using probability plot, figure 2 to figure 7, shows the Normal, Log-normal, Exponential, Wei bull and Gamma probability distribution respectively fit over the collected data using the probability plot of the data, the lines into the figures are the distribution lines which pass through the data points of the collected data. The distribution fit is to be considered fit the distribution when the distribution line pass through the data points, the much closer to the distribution line is much fitted distribution, so figure3 show that the distribution line is passing through the data points and first few data points and last are away from the distribution line, similarly the figure4 and figure 6, the distribution line is much closer to the collected data, while the figure5 shows exponential distribution and the distribution is not fitted over the collected data The figure 8 shows the standard Error in to the distribution fit, the 5% or less error shown in to the log normal and gamma distribution which is 0.0299142 and 2.23807 respectively, so the log normal and gamma distribution are the best fit for the data and among then the log normal shown much best fit for the data. After describing the all fitted distributions, for the fitted distribution the stochastic behavior of the stochastic traffic flow model must behavior like the distribution or it can be considered that by using the most fitted probability distribution is likely to apply over the traffic flow for the stochastic behavior of the traffic flow.

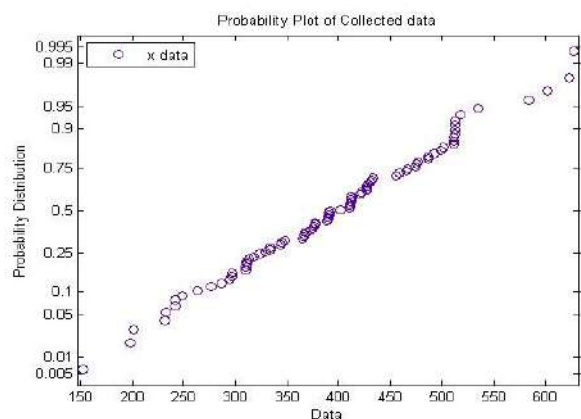


Figure 2. Probability plot of the collected data

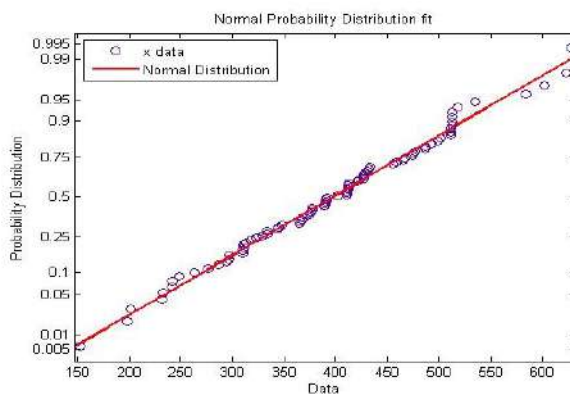


Figure 3. Normal probability distribution fit

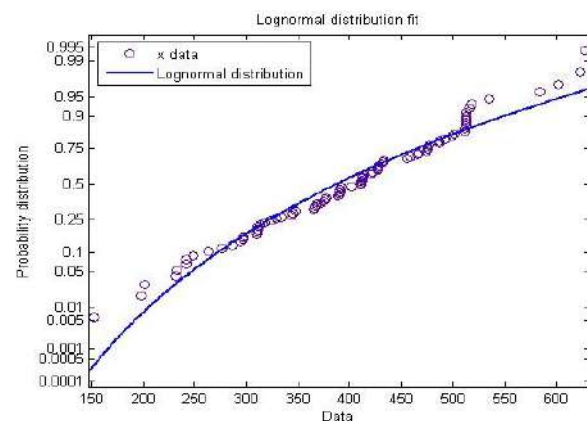


Figure 4. Lognormal probability distribution fit

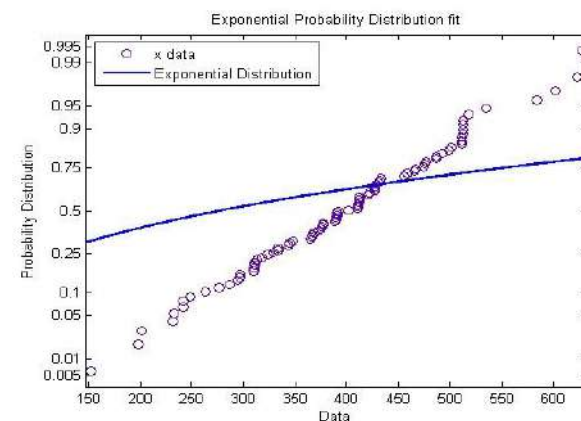


Figure 5. Exponential probability distribution fit

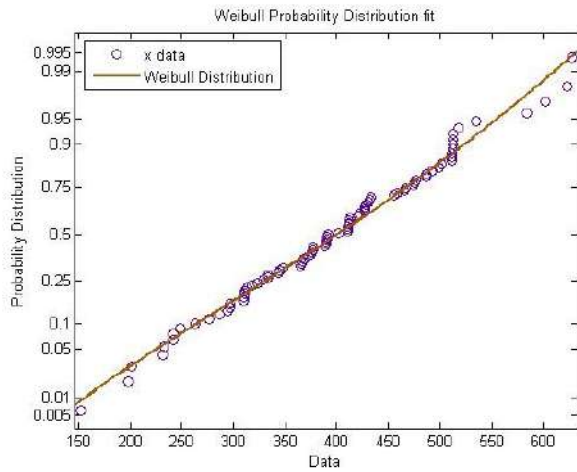


Figure 6. Weibull probability distribution fit

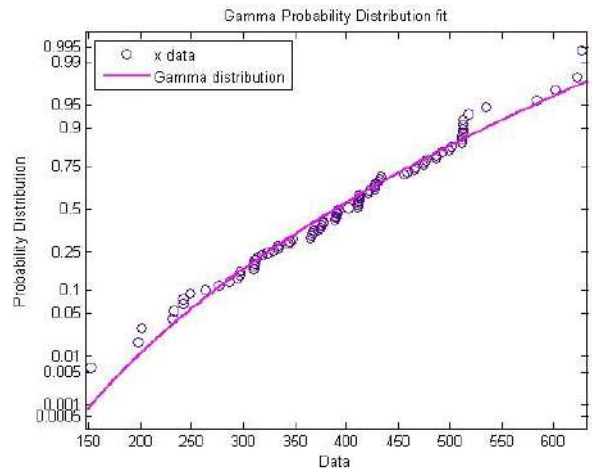


Figure 7. Gamma probability distribution fit

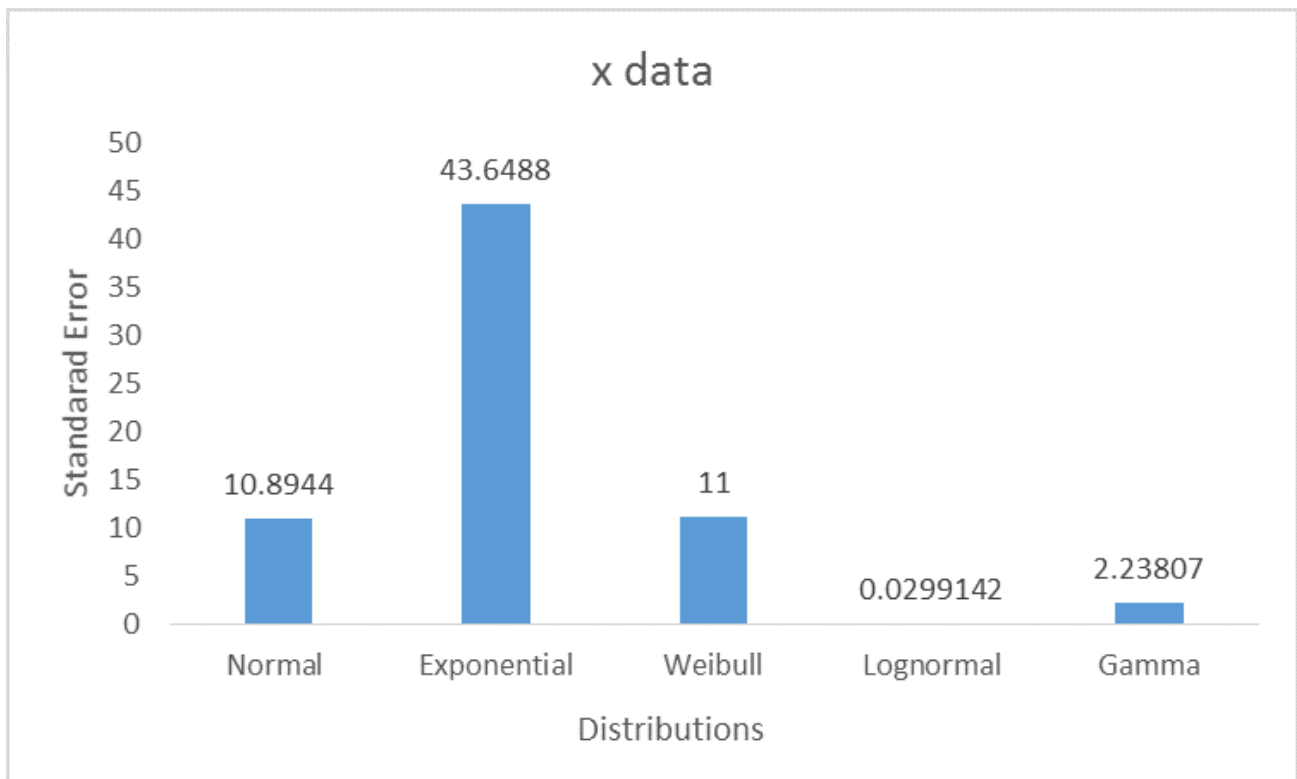


Figure 8. Standard Error Estimation in probability distribution fit

The data set is tested using Kolmogorov-Smirnov (K-S) test, Kolmogorov-Smirnov Modified test, and Anderson darling (AD) test is applied for Normal, lognormal, exponential, Weibull and gamma distribution and the result is shown in table 2, which shows the result of statistics, p-value, and decision at 95% of confidence level. Using the table of values from decision it's clear that the distribution which accept the all hypothesis test are Normal distribution, Weibull distribution and Gamma distribution while the Exponential distribution all the hypothesis test rejects and for Lognormal distribution the K-S and K-S modified test accept the hypothesis while A-D test reject the hypothesis test, concluding, the Normal, Weibull and Gamma distributions are best fit distribution for the collected data over the section of road, while among Normal, Weibull and Gamma distribution taking rank of p-values the normal distribution is the best fit over the data. Figure 8, shows the standard error into the distribution which is calculated which applied the distribution fit using the MATLAB distribution fit tool box, for each distribution using the distribution tool box in MATLAB gives the standard error into the distribution fit. Similarly the same is also calculated by using the formula $SE = \sigma\sqrt{n}$, where sigma is the standard deviation of the collected data and n is the length of collected data, as in figure 8, the maximum error is in exponential distribution, while for others distribution the normal distribution and Weibull distribution has the error more than allowable error which is greater than 5% while the lognormal and gamma distributions has the minimum standard error which is under the 5%.

Table 3 goodness-of-fit test for the collected data

Distribution	Parameter	Estimation	GOF Test	Statistics	p-value	Decision
Normal	Location mu	400.05	K-S test	0.05922	1	Can't reject Normal
	Scale sigma	99.848	K-S modified	0.05922	> 0.15	<i>Can't reject Normal</i>
			A-D test	0.31649	0.53398	<i>Can't reject Normal</i>
Log normal	Location mu	5.957	<i>K - Stest</i>	0.10783	0.26392	<i>Can't reject Lognormal</i>
	Scale sigma	0.274	<i>K - Smodified</i>	0.09592	> 0.15	<i>Can't reject Lognormal</i>
Exponential	Mean theta	0.003	<i>K - Stest</i>	0.05922	1	<i>Can't reject Exponential</i>
		0.003	<i>K - Smodified</i>	0.05922	> 0.15	<i>Can't reject Exponential</i>
Weibull	Scale alpha	0.003	<i>K - Stest</i>	0.05922	1	<i>Can't reject Weibull</i>
	Shape beta	0.003	<i>K - Smodified</i>	0.05922	> 0.15	<i>Can't reject Weibull</i>
Gamma	Shape alpha	0.003	<i>K - Stest</i>	0.05922	1	<i>Can't reject Gamma</i>
	Scale theta	0.003	<i>K - Smodified</i>	0.05922	> 0.15	<i>Can't reject Gamma</i>

4 CONCLUSION

This study is based over the determination of best probability distribution fit over the collected traffic flow data over the section of road of Nawabshah Pakistan. The probability distribution fit is used by distribution fit on MATLAB and the hypothesis test applied over the collected data, totally five distribution fit used namely Normal distribution, Lognormal distribution, Exponential distribution, Weibull distribution and Gamma distribution fit. Among these five distribution fit only three distribution fit are accept that are Normal, Weibull and Gamma distribution and from these fitted distributions the best fit is lognormal distribution. Three hypothesis test Kolmogorov-Smirnov (K-S) test, Kolmogorov-Smirnov Modified test, and Anderson darling (AD) applied over the collected data and by using 95%of confidence level, these hypothesis accept the test only for Normal, Weibull and Gamma distributions and using rank by p-values the most best fit distribution is lognormal and Gamma probability distribution. Hence concluding that the stochastic behavior of the collected data is like the lognormal distribution and gamma distribution. In future the same study can be carry out by incorporate the data with ramps on and off, by subdivided the traffic flow data in vehicles types, section of road with inflow and outflow data from different road section and over the boundaries of the road section by neglecting ramp in or out, by considering more than one path and collecting data for more than one path, by changing the different method of collecting data and many more.

5 Author Contributions

Mehboob Ali Jatoi¹ :Conceptualization, methodology, write-up, **Shakeel Ahmed Kamboh**:Draft write-up, **Oshaque Ali Abro**: mathematical editing, **Saeed Ahmed Rajput**:Draft review and paper formating, **Liaquat Ali Zardari**: Draft review.

6 Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

7 Funding Information

No Funding

8 Author Information

ORCID:

Mehboob Ali Jatoi¹: [0009-0009-8134-4262](https://orcid.org/0009-0009-8134-4262)

Shakeel Ahmed Kamboh: [0000-0002-3468-1663](https://orcid.org/0000-0002-3468-1663)

shaque Ali Abro: [0009-0005-9892-0888](https://orcid.org/0009-0005-9892-0888)

Saeed Ahmed Rajput: [0009-0007-3310-9140](https://orcid.org/0009-0007-3310-9140)

Liaquat Ali Zardari: [00009-0008-4659-279X](https://orcid.org/00009-0008-4659-279X)

References

- [1] Abul-Magd, A. [2007], 'Modeling highway-traffic headway distributions using superstatistics', *Physical Review E* **76**(5), 057101.
- [2] Alasmar, M., Clegg, R., Zakhleniuk, N. and Parisis, G. [2021], 'Internet traffic volumes are not gaussian—they are log-normal: An 18-year longitudinal study with implications for modelling and prediction', *IEEE/ACM Transactions on Networking* **29**(3), 1266–1279.
- [3] Ali, Y., Sharma, A., Haque, M. M., Zheng, Z. and Saifuzzaman, M. [2020], 'The impact of the connected environment on driving behavior and safety: A driving simulator study', *Accident Analysis & Prevention* **144**, 105643.
- [4] Alperovich, T. and Sopasakis, A. [2008], 'Stochastic description of traffic flow', *Journal of Statistical Physics* **133**, 1083–1105.
- [5] Antoniou, I., Ivanov, V. V., Ivanov, V. V. and Zrelov, P. [2002], 'On the log-normal distribution of network traffic', *Physica D: Nonlinear Phenomena* **167**(1-2), 72–85.
- [6] Berry, D. S. and Belmont, D. M. [1951], Distribution of vehicle speeds and travel times, in 'Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability', Vol. 2, University of California Press, pp. 589–603.
- [7] Bramich, D. M., Menendez, M. and Ambühl, L. [2022], 'Fitting empirical fundamental diagrams of road traffic: A comprehensive review and comparison of models using an extensive data set', *IEEE Transactions on Intelligent Transportation Systems* **23**(9), 14104–14127.
- [8] Chakravarti, I. M., Laha, R. G. and Roy, J. [1967], 'Handbook of methods of applied statistics', *Wiley Series in Probability and Mathematical Statistics (USA) eng* .
- [9] Cheng, Q., Lin, Y., Zhou, X. S. and Liu, Z. [2024], 'Analytical formulation for explaining the variations in traffic states: A fundamental diagram modeling perspective with stochastic parameters', *European Journal of Operational Research* **312**(1), 182–197.
- [10] Chiabaut, N. and Faitout, R. [2021], 'Traffic congestion and travel time prediction based on historical congestion maps and identification of consensual days', *Transportation Research Part C: Emerging Technologies* **124**, 102920.
- [11] Das, K. R. and Imon, A. [2016], 'A brief review of tests for normality', *American Journal of Theoretical and Applied Statistics* **5**(1), 5–12.
- [12] Dubey, S. K., Ponnu, B. and Arkatkar, S. S. [2013], 'Time gap modeling using mixture distributions under mixed traffic conditions', *Journal of Transportation Systems Engineering and Information Technology* **13**(3), 91–98.

- [13] Fang, Z., Zhu, S., Fu, X., Liu, F., Huang, H. and Tang, J. [2022], 'Multivariate analysis of traffic flow using copula-based model at an isolated road intersection', *Physica A: Statistical Mechanics and its Applications* **599**, 127431.
- [14] Fras, M., Mohorko, J. and Čučej, Ž. [2012], 'Modeling of statistical data sources based on measured network traffic', *Simulation* **88**(10), 1216–1232.
- [15] Ghosh, S., Roy, M. K. and Biswas, S. C. [2016], 'Determination of the best fit probability distribution for monthly rainfall data in bangladesh', *American Journal of Mathematics and Statistics* **6**(4), 170–174.
- [16] Guessous, Y., Aron, M., Bhourri, N. and Cohen, S. [2014], 'Estimating travel time distribution under different traffic conditions', *Transportation Research Procedia* **3**, 339–348.
- [17] Handscombe, J. and Yu, H. Q. [2019], 'Low-cost and data anonymised city traffic flow data collection to support intelligent traffic system', *Sensors* **19**(2), 347.
- [18] Khakifirooz, M., Tercero-Gómez, V. and Woodall, W. H. [2021], 'The role of the normal distribution in statistical process monitoring', *Quality Engineering* **33**(3), 497–510.
- [19] Kosun, C. and Ozdemir, S. [2016], 'A superstatistical model of vehicular traffic flow', *Physica A: statistical mechanics and its applications* **444**, 466–475.
- [20] Laurikkala, M. [2009], 'Goodness-of-fit tests and heavy-tailed distributions in network traffic data analysis'.
- [21] Leduc, G. et al. [2008], 'Road traffic data: Collection methods and applications', *Working Papers on Energy, Transport and Climate Change* **1**(55), 1–55.
- [22] Li, L., Jiang, R., He, Z., Chen, X. M. and Zhou, X. [2020], 'Trajectory data-based traffic flow studies: A revisit', *Transportation Research Part C: Emerging Technologies* **114**, 225–240.
- [23] Li, M., Fang, L., Jia, W. and Guo, J. [2022], 'Traffic condition uncertainty quantification under nonnormal distributions', *Journal of transportation engineering, Part A: Systems* **148**(10), 04022086.
- [24] Mondal, S. and Gupta, A. [2021], 'Speed distribution for interrupted flow facility under mixed traffic', *Physica A: Statistical Mechanics and Its Applications* **570**, 125798.
- [25] Monti, G. S., Mateu Figueras, G., Ortego Martínez, M. I., Pawlowsky Glahn, V. and Egozcue Rubí, J. J. [2017], Modified kolmogorov-smirnov test of goodness of fit, in 'The 7th International Workshop on Compositional Data Analysis Proceedings book', pp. 152–158.
- [26] Pun, L., Zhao, P. and Liu, X. [2019], 'A multiple regression approach for traffic flow estimation', *IEEE access* **7**, 35998–36009.
- [27] Sánchez, E., González-Navarrete, M. and Caamaño-Carrillo, C. [2021], 'Bivariate superstatistics: An application to statistical plasma physics', *The European Physical Journal B* **94**, 1–7.
- [28] Stephens, M. A. [1974], 'Edf statistics for goodness of fit and some comparisons', *Journal of the American statistical Association* **69**(347), 730–737.

- [29] Ximenes, P. d. S. M. P., Silva, A. S. A. d., Ashkar, F. and Stosic, T. [2021], 'Best-fit probability distribution models for monthly rainfall of northeastern brazil', *Water Science and Technology* **84**(6), 1541–1556.
- [30] Yang, Y., Fan, Y. and Royset, J. O. [2019], 'Estimating probability distributions of travel demand on a congested network', *Transportation Research Part B: Methodological* **122**, 265–286.
- [31] Zang, Z., Xu, X., Qu, K., Chen, R. and Chen, A. [2022], 'Travel time reliability in transportation networks: A review of methodological developments', *Transportation Research Part C: Emerging Technologies* **143**, 103866.
- [32] Zefreh, M. M. and Török, A. [2020], 'Distribution of traffic speed in different traffic conditions: An empirical study in budapest', *Transport* **35**(1), 68–86.