[Physical Education, Health and Social Sciences](https://journal-of-social-education.org)

https://journal-of-social-education.org

E-ISSN: 2958-5996
P-ISSN: 2958-5988

Crime Patterns in Pakistan: Analyzing Trends and Crime Rates Using Statistical Distributions

Ramla Shah[1], Sidra Younas[2], Dr. Maryam Sadiq[3]

[1, 2, 3] Department of Statistics, University of Azad Jammu & Kashmir,
ramlashah193@gmail.com, sidra.younas@ajku.edu.pk, hussainulahmad@gmail.com

Abstract

Crime remains a significant social challenge, impacting communities and governance worldwide. This study analyzes crime trends in Pakistan from 1997 to 2019, focusing on identifying the best-fitting probability distributions for different crime types and provincial crime rates. Using secondary data from the Pakistan Bureau of Statistics, we categorize offenses and apply statistical models to determine their distributional properties. We evaluate four probability distributions—Frechet, Log-Normal, Log-Logistic, and Log-Gamma—and estimate their parameters using Maximum Likelihood Estimation (MLE) and Bayesian Estimation (BE). The Run Test checks the randomness of crime occurrences, while the Mann-Whitney U Test assesses distributional identicalness across provinces. The Kolmogorov-Smirnov Test determines the best fit for each crime type. Findings suggest that Frechet and Log-Normal distributions best describe crimes such as murder, robbery, and theft, while Log-Gamma and Log-Logistic fit more dispersed crime categories. Punjab and Sindh report the highest crime frequencies, with skewed distributions. These insights enhance crime forecasting, aiding policymakers in data-driven crime prevention strategies. This study contributes to criminology by applying probabilistic models to crime analysis. Future research may integrate socio-economic variables to improve predictive accuracy and develop more effective law enforcement policies.

Keywords: Probability Distributions, Crime Trends, Statistical Distributions, Maximum Likelihood Estimation, Bayesian Estimation.

Introduction

Crime is a crucial topic to inspect and study and becomes a major social problem in all parts of the world. However, it concerns us when it touches our own lives. The modern perception of crime is the breach of individual rights. According to Black's Law Dictionary, "A crime is an act committed or omitted in violation of a public law, either forbidding or commanding it; A breach or violation of some public right or duty due to a whole community, considered as a community in its social aggregate capacity, as distinguished from a civil injury." A mind set of crimes arises when people engage in unethical activities with impudence and expected to be ignored by the law. The consequences of crimes increased with the acceleration of the progression of economic upturn, modernization, social status mobility, increased unemployment, economic diversification of society, normative chaos and others. During the last few decades, crimes in Asia were increased significantly. Pakistan is the worst hit country from terrorism; hundreds of crimes take place on a daily basis. Some of these crimes include robbery, vehicle snatching, kidnapping, murder, rape, looting of pedestrians and paraphilia. This had led to a fearful society with masses losing faith in government and law enforcement agencies. Crime records do not envelope all factual offenses; thus, none of the institutions provide full data. This set of circumstances allows two types of crimes to be explained, crimes which are reported and exposed by authorities and concealed crimes. Wallace (2017) assessed the impact of organized crime on SIDS (sudden infant death syndrome) in the Caribbean. He indicated negative social, political, communal and economic, governance and governability impacts on Caribbean societies as a result of organized crime were indicated by using archival research that is secondary data and self administered questionnaires. Akinrefon (2016) curbed the menace of crimes in Adamawa state Nigeria. Crime rates were to increase over the years and age group of 16-35 was most involved in criminal acts. Khan *et al*., (2015) mentioned that crimes were related to the level of education attained and social and economic background of an individual. There was a significant negative relationship between crime rates and higher education, but a positive relationship was found between crime rates and poverty. Fitterer and Nelson (2015) modeled the relationship between alcohol consumption and crime. The relation of alcohol exposure and the rate of crimes were positive and showed by using multiple spatial units to capture spatial effects. Khalid *et al*., (2015) explained that the majority of crimes were found in commercial and densely populated areas. He conducted the study in the Faisalabad city of Pakistan by using the geographic information system to detect the hot spots of street crimes. According to statistical point of view, it is important to study the distribution of crimes. This study was undertaken to highlight the total crime rates in Pakistan since 1997 to 2018 and to highlight the distributions of crimes by types of different sites of Pakistan. Efficient parameter estimation method and appropriate distribution of crimes during the year 2011-2016 was determined using the Kolmogorov Smirnov test.

Methodology

Data for this research provides crime conditions in Pakistan from 1997 to 2018. The data used is secondary data extracted from Pakistan Bureau of Statistics. Crime data for six years (2011-16) is detailed with crime types and regions included. The R and Mathematica software were used to carry out the analysis. Crude crime rates are obtained for year 1997-18. These are taken by dividing the number of criminal cases within a specific year by the total midyear population of that year.

$$Crude\ rate = \frac{number\ of\ criminal\ cases\ in\ a\ given\ year}{the\ total\ midyear\ population\ of\ the\ given\ year} \times 1000$$

Parametric Distribution Fitting

Parametric distributional forms are helpful for describing data in parsimonious manner and for developing conceptual models for felonious events. In this study we will find the appropriate distribution(s) for crimes data. We will separately analyze crimes by types and crimes by site. Crime types under study are murder, attempt to murder, kidnapping, robbery, theft and miscellaneous crimes. Crime sites considered are Punjab, Sindh, Khyber Pakhtunkhwah, Balochistan, Islamabad, Gilgit Baltistan and Azad Kashmir. Crime counts are analyzed by fitting the four different parametric distributions. Our proposed distributions for this purpose are Frechet, Log-normal, Log-logistic and Log-gamma distributions. The assumptions of randomness and identicalness are checked by using Run and Mann-Whitney U test, respectively. Then the parameters of distributions are estimated by using Maximum likelihood estimation and Bayesian estimation methods.

Run Test

Run test is a nonparametric test. It is used to check whether data under study is random or not. The test statistic is as

$$Z = \frac{R - \underline{R}}{S_R}$$

Where R= number of runs, $\underline{R}$= expected number of runs and S_R = standard deviation of number of runs. While,

$$\underline{R} = \frac{2n_1 n_2}{n_1 + n_2} + 1$$

$$S_R = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

If p-value is greater than level of significance, the null hypothesis will be accepted that the data is random otherwise hypothesis will be rejected.

Mann Whitney U Test

The Mann Whitney U test is a nonparametric test based on ranks. The test is used to determine that the two independent samples follow the identical distribution. The test statistic is

$$U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

As, U=Mann Whitney U test, n_1 *and* n_2 *are sample sizes and* $R =$ rank of the sample. It depends on the p-values of the test statistic whether to reject or accept the hypothesis. If p-values are greater than the level of significance, then we accept the hypothesis otherwise reject.

Frechet Distribution

Let $X_{1,} X_2, \ldots, X_n$ denote a random sample of size 'n' from FD having two parameters. The Probability Density Function (PDF) of FD is,

$$f(x,\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1} e^{-\left(\frac{\beta}{x}\right)^{\alpha}}, \qquad x > 0, \alpha, \beta > 0$$

Where, α and β are shape and scale parameters respectively. The Cumulative Density Function (CDF) of FD,

$$F(x) = e^{-\left(\frac{\beta}{x}\right)^{\alpha}}$$

ML Estimators of FD

Let $X_{1,} X_2, \ldots, X_n$ denote a random sample of size 'n' from FD, then the likelihood of function is

$$L = \alpha^n \beta^{n\alpha} \prod_{i=1}^{n} x_i^{-(\alpha+1)} e^{-\sum_{i=1}^{n} \left(\frac{\beta}{x_i}\right)^{\alpha}}$$

The log likelihood function is

$$Log\ L = nlog(\alpha) - n\alpha\ log(\beta) - (\alpha + 1)\sum_{i=1}^{n} logx_i - \sum_{i=1}^{n} \left(\frac{\beta}{x_i}\right)^{\alpha}$$

By partially differentiating equation (3.6) with respect to α and β and putting equal to 0, we get the following normal equations

$$\frac{\partial\ log\ log\ L}{\partial\alpha} = \frac{n}{\alpha} + nlog\ \beta - \sum_{i=1}^{n} logx_i - \sum_{i=1}^{n}\left[\left(\frac{\beta}{x_i}\right)^{\alpha} log\ log\left(\frac{\beta}{x_i}\right)\right] = 0 \quad (a)$$

$$\frac{\partial logL}{\partial\beta} = \frac{n\alpha}{\beta} - \frac{\alpha}{\beta}\sum_{i=1}^{n}\left[\left(\frac{\beta}{x_i}\right)^{\alpha}\right] = 0 \qquad (b)$$

Equation (a) and (b) are not in closed form. Which can be solved by using basic fibroblast growth factor (BFGF) method.

BE of FD

The reference prior has been used to get BE (Abbas and Tang, 2015). Which is

$$\pi_R = \frac{1}{\alpha\beta}$$

The joint posterior distribution for α, β is

$$\pi(x) = \frac{\alpha^{n-1}\,\beta^{n\alpha-1}\prod_{i=1}^{n} x_i^{-(\alpha+1)}\, exp\,exp\left[-\sum_{i=1}^{n}\left(\frac{\beta}{x_i}\right)^{\alpha}\right]}{\int_0^{\infty}\int_0^{\infty}\alpha^{n-1}\beta^{n\alpha-1}\prod_{i=1}^{n} x_i^{-(\alpha+1)}\, exp\left[-\sum_{i=1}^{n}\left(\frac{\beta}{x_i}\right)^{\alpha}\right] d\alpha d\beta} \qquad (c)$$

By using equation (c) here we use a Laplace approximation for Bayesian estimates α and β.

Lognormal Distribution (LND)

Let $X_1, X_2, \ldots, X_n$ denotes a random sample of size 'n' from LND. The PDF of lognormal distribution is

$$f(x,\mu,\sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}}\, e^{-\frac{(logx-\mu)^2}{2\sigma^2}}, \qquad x > 0, \sigma^2 > 0$$

Where, μ and σ^2 are location and scale parameter respectively.

The CDF of LND is

$$F(x) = \int_{-\infty}^{x} \frac{1}{x\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left[\frac{loglog\,(x)-\mu}{\sigma}\right]^2} dx \qquad x > 0, \sigma > 0$$

ML Estimators of LND

Likelihood function is

$$L = (2\pi\sigma^2 x^2)^{-\frac{n}{2}}\, e^{-\frac{\sum_{i=1}^{n}(logx-\mu)^2}{2\sigma^2}}$$

By taking log, we get the following equation

$$logL = -\frac{n}{2}\, log\, log\,(2\pi\sigma^2 x^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(logx_i-\mu)^2 \qquad (d)$$

Partial differentiate equation (d) with respect to μ and put equal to zero

$$\frac{\partial logL}{\partial\mu} = \frac{1}{\sigma^2}\left(n\mu - \sum_{i=1}^{n} logx\right) = 0$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} logx}{n} = \underline{X}$$

Partial differentiate equation (d) with respect to σ^2 and put equal to zero

$$\frac{\partial logL}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(logx-\hat{\mu})^2 = 0$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(logx-\hat{\mu})^2}{n}$$

BE of LND

To develop the BE, the Jeffrey's prior is used as:

$$\pi_{(x)} = \frac{1}{\sigma}$$

The Posterior distribution is

$$\pi_{(x)} \propto \frac{(2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}}{\prod_{i=1}^{n} x_i}\, e^{-\frac{\sum_{i=1}^{n}(logx-\mu)}{2\sigma^2}} \cdot \frac{1}{\sigma}$$

$$\pi_{(x)} = ke^{-\frac{1}{2}\left[\frac{\mu-\underline{X}}{\sqrt{\frac{\sigma^2}{n}}}\right]^2} \qquad (e)$$

The value of 'k' is

$$k = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sqrt{\sigma^2}}$$

Putt values of 'k' in equation (e)

$$\pi_{(x)} = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sqrt{\sigma^2}}\; e^{-\frac{1}{2}\left[\frac{\mu - \underline{X}}{\sqrt{\frac{\sigma^2}{n}}}\right]^2}$$

$$E(x) = \int_{-\infty}^{\infty} \mu.\frac{\sqrt{n}}{\sqrt{2\pi}\,\sqrt{\sigma^2}}\; e^{-\frac{1}{2}\left[\frac{\mu - \underline{X}}{\sqrt{\frac{\sigma^2}{n}}}\right]^2} d\mu$$

$$= \frac{\sqrt{n}}{\sqrt{2\pi}\,\sqrt{\sigma^2}}\int_{-\infty}^{\infty} \mu.\, e^{-\frac{1}{2}\left[\frac{\mu - \underline{X}}{\sqrt{\frac{\sigma^2}{n}}}\right]^2} d\mu$$

Let,

$$\left[\frac{\mu - \underline{X}}{\sqrt{\frac{\sigma^2}{n}}}\right] = z \Longrightarrow \mu = \underline{X} + \sqrt{\frac{\sigma^2}{n}}z \;\Longrightarrow\; d\mu = \underline{X} + \sqrt{\frac{\sigma^2}{n}}dz$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \underline{X}\, e^{-\left(\frac{z^2}{2}\right)} dz + \int_{-\infty}^{\infty} \sqrt{\frac{\sigma^2}{n}} z\, e^{-\left(\frac{z^2}{2}\right)} dz$$

$$E(x) = \frac{1}{\sqrt{2\pi}}\,\underline{X}\int_{-\infty}^{\infty} e^{-\left(\frac{z^2}{2}\right)} dz$$

$$E(x) = \sum_{i=1}^{n} \frac{logx}{n} = \underline{X}$$

Loglogistic Distribution

The PDF of LLD is

$$f\left(x_{\,;}\,\alpha,\beta\right) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right)^2}\;; \qquad \alpha,\beta > 0 \qquad (3.17)$$

Where α and β is scale and shape parameter respectively.

The CDF of LLD is

$$F\left(x_{\,;}\,\alpha,\beta\right) = \frac{1}{1+\left(\frac{x}{\alpha}\right)^{-\beta}} \qquad (3.18)$$

ML Estimators of LLD

The likelihood function of equation (3.17)

$$L = \beta^n\alpha^{-n\beta}\prod_{i=1}^{n} {x_i}^{\beta-1}\prod_{i=1}^{n}\left[1+\left(\frac{x_i}{\alpha}\right)^{\beta}\right]^{-2} \qquad (3.19)$$

Taking log on equation (3.19) we get,

$$logL = nlog(\beta) - n\beta log\,(\alpha) + (\beta-1)\sum_{i=1}^{n} log\,(x_i) - 2\sum_{i=1}^{n} log\left[1+\left(\frac{x_i}{\alpha}\right)^{\beta}\right]$$

By differentiating the above equation with respect to α,β and put equal to zero, we get

$$\frac{\partial logL}{\partial \alpha} = -\frac{n\beta}{\alpha} + \frac{2\beta}{\alpha}\sum_{i=1}^{n} \left(\frac{x_i}{\alpha}\right)^{\beta}\left[1+\left(\frac{x_i}{\alpha}\right)^{\beta}\right]^{-1} = 0 \qquad (3.20)$$

$$\frac{\partial logL}{\partial \beta} = \frac{n}{\beta} - nlog(\alpha) + \sum_{i=1}^{n} log(x_i) - 2\sum_{i=1}^{n} \left(\frac{x_i}{\alpha}\right)^{\beta} log\left(\frac{x_i}{\alpha}\right)\left[1+\left(\frac{x_i}{\alpha}\right)^{\beta}\right]^{-1} = 0 \qquad (3.21)$$

Equation (3.20) and (3.21) are not in closed form. So an iterative method is used to obtain the estimates of the parameters.

BE of LLD

The reference prior has been used to get BE of parameters for LLD, (Abbas and Tang, 2016). The joint posterior distribution of α and β is

$$\pi_{(x)} = \frac{\alpha^{n-1}\beta^{n\alpha-1}\prod_{i=1}^{n} x_i^{\beta-1}\prod_{i=1}^{n}\left[1+\left(\frac{x_i}{\alpha}\right)^{-\beta}\right]^{-2}}{\int_0^{\infty}\int_0^{\infty} \alpha^{-n\beta-1}\beta^{n-1}\prod_{i=1}^{n} x_i^{\beta-1}\prod_{i=1}^{n}\left[1+\left(\frac{x_i}{\alpha}\right)^{-\beta}\right]^{-2} d\alpha d\beta} \qquad (3.22)$$

An iterative method is used to obtain the estimates of the parameters using equation (3.22).

Log-Gamma Distribution (LGD)

A random variable X is distributed as LGD if its natural log is gamma distributed with α and β as scale and shape parameter respectively. PDF of LGD is

$$f(x\,;\alpha,\beta) = \frac{e^{\beta x}\, e^{-\left(\frac{x}{\alpha}\right)}}{\alpha^{\beta}\,\Gamma(\beta)} \qquad (3.23)$$

The CDF of this distribution is not in closed form.

ML Estimators of LGD

$$L = \frac{e^{\beta\sum_{i=1}^{n} x}\, e^{-\sum_{i=1}^{n}\left(\frac{x}{\alpha}\right)}}{\alpha^{n\beta}\,(\Gamma(\beta))^{n}} \qquad (3.24)$$

The log likelihood function is

$$logL = \frac{\beta\sum_{i=1}^{n} x - \sum_{i=1}^{n}\left(\frac{x}{\alpha}\right)}{n\beta log\alpha + nlog\Gamma(\beta)} \qquad (3.25)$$

By differentiate partially equation (3.25) with respect to α, β and put equal to zero we get

$$\frac{\partial logL}{\partial \alpha} = \frac{\sum_{i=1}^{n} x}{\alpha} - n\beta log\alpha - nlog\Gamma(\beta) \qquad (3.26)$$

$$\frac{\partial logL}{\partial \beta} = \sum_{i=1}^{n} x - nlog\alpha - \frac{\partial}{\partial \beta}\left(nln\Gamma(\beta)\right) \qquad (3.27)$$

Equation (3.26) and (3.27) are not in close form. Which can be solved by iterative method.

BE of LGD

Let $X_{1,}\, X_2, \dots, X_n$ be a random sample of size 'n' then the joint posterior distribution for α, β is

$$\pi_{(x)} = \frac{e^{\beta\sum_{i=1}^{n} x}\, e^{-\sum_{i=1}^{n}\left(\frac{x}{\alpha}\right)}}{\alpha^{n\beta}\,(\Gamma(\beta))^{n}} \cdot \frac{1}{\alpha\beta} \qquad (3.28)$$

$$\propto \frac{e^{\beta\sum_{i=1}^{n} x - \sum_{i=1}^{n}\left(\frac{x}{\alpha}\right)}}{\alpha^{n\beta+1}\,(\Gamma(\beta))^{n}} \cdot \frac{1}{\beta}$$

$$\propto e^{\sum_{i=1}^{n} x(\alpha)}\,(\Gamma(\beta))^{-n}.\,\beta^{-1}\,\alpha^{-(n\beta+1)}$$

$$= \int_{-\infty}^{\infty} \alpha^{-(n\beta+1)} e^{-\sum_{i=1}^{n} \left(\frac{x}{\alpha}\right)} d\alpha \qquad (3.29)$$

Which can be solved by iterative methods.

Goodness of Fit test

Kolmogorov Smirnov goodness of fit test is applied at the 5 percent level of significance to find out the best fit distribution.

Kolmogorov Smirnov Test

Suppose we have a random sample $X_1, X_2, \ldots, X_n$ from some distribution having CDF F (x_i).
The test statistic is D= max (D^+, D^-)
Where

$$D^+ = max\left(\frac{i}{n} F(x_i)\right)$$

$$D^- = \left(F(x_i) - \frac{i-1}{n}\right)$$

Results and Discussions

Table 4.1: Crime Rates from 1997 to 2018

Years	Total crimes recorded	Total Population	Crime Rate Per 1000	Crime Rate
1997	370254	129086987	2.86825193	0.002868
1998	431792	132352279	3.26244477	0.003262
1999	417799	135405584	3.08553745	0.003086
2000	388859	138523285	2.80717426	0.002807
2001	378301	141601437	2.67159012	0.002672
2002	399568	144654143	2.76222991	0.002762
2003	400680	147703401	2.71273374	0.002713
2004	440578	150780300	2.92198649	0.002922
2005	453264	153909667	2.9450002	0.002945
2006	524137	157093993	3.33645476	0.003336
2007	524208	160332974	3.26949589	0.003269
2008	577420	163644603	3.52850011	0.003529
2009	616227	167049580	3.68888686	0.003689
2010	652383	170560182	3.82494315	0.003825
2011	673750	174184265	3.86803022	0.003868
2012	645647	177911533	3.62903399	0.003629
2013	634404	181712595	3.49124946	0.003491
2014	627127	185546257	3.37989572	0.00338
2015	633299	189380513	3.34405578	0.003344
2016	677554	193203476	3.50694519	0.003507
2017	683925	207774520	3.26244477	0.003292
2018	703481	200813818	3.50315037	0.003503

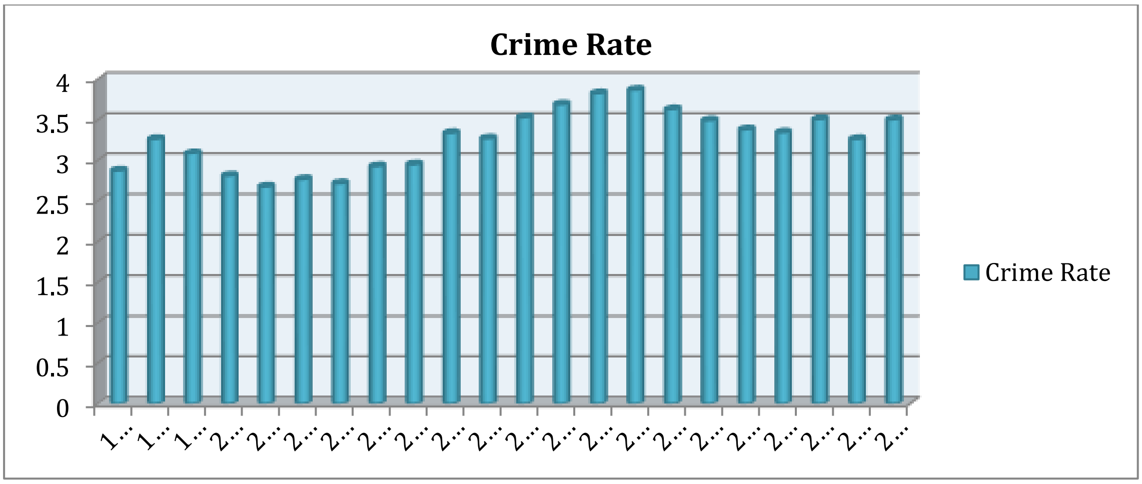

Table 4.2: Descriptive Statistics for crime data

Crimes	**Mean**	**C.V.**	**Variance**	**Min.**	**Max.**	**Skewness**	**Kurtosis**
Murder	1519	130.4806	3929294	2.0	6459	1.1282	3.0269
Attempt to murder	1700	134.3968	5219551	4.0	7772	1.3503	3.7859
Kidnapping	2407	194.9937	22023583	1.0	15699	2.1133	5.7814
Robbery	2111	200.5692	17925467	1.0	15316	2.0578	5.7697
Theft	4627	226.4946	10981654	48.0	36400	2.2884	6.3902
Other crimes	65232	156.7205	104514081	959.0	328610	1.5743	4.0783

Note: Min. = Minimum, Max. = Maximum, and C.V. = Coefficient of Variation.

Table 4.3: Descriptive Statistics for sites wise crime data

Sites	**Mean**	**C.V.**	**Variance**	**Min.**	**Max.**	**Skewness**	**Kurtosis**
Punjab	44176	210.9777	8686655931	825.00	328610	2.4481	7.0902
Sindh	8449	206.8673	305504162	419.00	60777	2.4647	7.1329
KPK	16636	267.2644	1976936535	50.00	163403	2.5083	7.3812
Balochistan	951	209.7622	3979070	30.00	7182	2.4826	7.3162
Islamabad	837	222.1871	3457130	12.00	6564	2.4627	7.1863
G.B	168	198.0656	110820.3	2.00	1226	2.4592	7.2616
AJK	653	226.8399	2193383	3.00	5135	2.4356	6.9914

Table 4.4: Mann Whitney statistics for different crimes

S.No	Crimes	U	P-values
1	Murder	244.500	0.370
2	Attempt to murder	267.000	0.665
3	Kidnapping	267.500	0.672
5	Robbery	281.000	0.885
6	Theft	243.000	0.353
7	Other crimes	273.000	0.757

Table 4.5: Mann Whitney statistics for sites wise crimes

S.No	Sites	U	P-values
1	Punjab	299.000	0.257
2	Sindh	244.000	0.037*
3	KPK	333.000	0.586
4	Balochistan	320.500	0.447
5	Islamabad	361.000	0.952
6	G.B	294.500	0.226
7	AJK	321.500	0.457

		MURDER		
Methods	$\hat{\alpha}$	$\hat{\beta}$	**KS**	**P-values**
ML	0.7346	333.6872	0.1496	0.2328
BE	0.7233	294.4123	0.1570	0.1873
		ATTEMPT TO MURDER		
ML	0.7981	430.6572	0.1508	0.2252
BE	0.7857	388.1055	0.1594	0.1746
		KIDNAPPING		
ML	0.6839	288.1821	0.0895	0.8366
BE	0.6727	250.3940	0.0998	0.7253
		ROBERRY		
ML	0.6869	166.1555	0.1498	0.2316
BE	0.6770	145.0378	0.1281	0.4105
		THEFT		
ML	0.9834	545.9570	0.1042	0.6363
BE	0.9704	512.6183	0.1044	0.6339
		OTHER CRIMES		
ML	1.1151	8.3376	0.2157	0.0194*
BE	1.0161	8.2403	0.1960	0.0432*

Parameter Estimation of Candidate Distribution

For the analysis of the crime data, we estimate the parameters of the selected distribution, such as FD, LLD, LND and LGD.

For Crimes Data

Tables 4.5, 4.6, 4.7, 4.8 showed the estimated parameters for FD, LLD, LND and LGD. Where $\hat{\alpha}$ is shape and $\hat{\beta}$ is scale parameters respectively.

Table 4.7: Estimates of Parameters for LLD

Table 4.6. Estimates of Parameters for FD

		Murder		
Methods	$\hat{\alpha}$	$\hat{\beta}$	**KS**	**P-values**
ML	0.4273	92.9343	0.1506	0.2262
BE	0.4275	82.6895	0.1682	0.1324
		Attempt to Murder		
ML	0.4451	127.2652	0.1760	0.1022
BE	0.4454	114.2875	0.1935	0.0549
		Kidnapping		
ML	0.4002	75.8371	0.1424	0.2848
BE	0.4007	66.3787	0.1606	0.1683
		Roberry		
ML	0.4184	50.8630	0.1384	0.3171
BE	0.4185	45.0051	0.1558	0.1946
		Theft		
ML	0.7118	288.2689	0.0867	0.8327
BE	0.7071	277.3178	0.0869	0.8305
		Other Crimes		
ML	0.6299	4997.6013	0.1750	0.0935
BE	0.6263	4752.5454	0.1790	0.0812

Table 4.8: Estimates of Parameters for LND.

MURDER				
Methods	$\hat{\mu}$	$\hat{\sigma}^2$	**KS**	**P-values**
ML	5.7159	2.2882	0.15401	0.18465
Attempt to Murder				
ML	5.9662	2.1497	0.15452	0.18181
Kidnapping				
ML	5.6067	2.5077	0.0875	0.82478
Roberry				
ML	5.1822	2.5308	0.16536	0.12916
Theft				
ML	6.4983	1.833	0.14636	0.23155
Other Crimes				
ML	9.4552	1.9655	0.24023	0.00637*

Table 4.9: Estimates of Parameters for LGD.

Murder				
Methods	$\hat{\alpha}$	$\hat{\beta}$	**KS**	**P-values**
ML	6.1099	0.93552	0.16173	0.14522
Attempt to Murder				
ML	7.5419	0.79108	0.16409	0.13457
Kidnapping				
ML	4.8945	1.1455	0.14788	0.22162
Roberry				
ML	4.1055	1.2622	0.14607	0.2335
Theft				
ML	12.306	0.52805	0.10893	0.58139
Other Crimes				
ML	22.659	0.41728	0.21423	0.02061*

Sites Wise Crimes Analysis

For this analysis, the main objective is to estimate the parameters of candidate distributions. Two methods of estimation named as ML and BE methods are used in the present study. P-values will be approximate in the presence of ties.

Table 4.10: Estimates of Parameters for FD.

Overall Crimes				
Methods	$\hat{\alpha}$	$\hat{\beta}$	**KS**	**P-values**
ML	3.6520	59.6120	0.9999	2.22e-16*
BE	3.5940	58.6310	0.9999	2.22e-16*
PUNJAB				
ML	0.9008	7058.0700	0.1631	0.1008
BE	0.8960	6913.8875	0.1588	0.1177
SINDH				
ML	1.0683	1675.1981	0.1770	0.0597
BE	1.0611	1652.2182	0.1752	0.0640
KPK				
ML	0.5948	370.9754	0.1282	0.3377
BE	0.5922	352.6566	0.1296	0.3249
BALOCHISTAN				
ML	1.0012	162.3114	0.0944	0.7222

BE	0.9949	159.6459	0.0955	0.7089
		ISIAMBAD		
ML	0.7942	91.3296	0.0985	0.6710
BE	0.7896	88.9489	0.1061	0.5770
		G.B		
ML	0.7612	26.3235	0.1219	0.3982
BE	0.7581	25.5474	0.1181	0.4384
		AJK		
ML	0.6487	61.7406	0.1808	0.0625
BE	0.6470	59.0388	0.1744	0.0795

Table 4.11: Estimates of Parameters for LLD

		Overall Crimes		
Methods	$\hat{\alpha}$	$\hat{\beta}$	**KS**	**P-values**
ML	1.5379	16241.5845	0.2276	0.0552
BE	1.5170	15910.3361	0.2213	0.0671
		PUNJAB		
ML	1.4324	11195.8438	0.1590	0.1166
BE	1.4139	10922.4993	0.1523	0.1471
		SINDH		
ML	1.6494	2480.6489	0.1605	0.1107
BE	1.6271	2434.7963	0.1564	0.1280
		KPK		
ML	0.8408	820.0747	0.1265	0.3532
BE	0.8304	758.9669	0.1176	0.4441
		BALOCHISTAN		
ML	1.4501	250.1583	0.1006	0.6453
BE	1.4316	244.1968	0.1004	0.6483
		ISIAMBAD		
ML	1.1830	162.9619	0.0959	0.7039
BE	1.1668	156.9254	0.0957	0.7058
		G.B		
ML	1.2839	48.9149	0.0967	0.6940
BE	1.2669	47.3229	0.0912	0.7600
		AJK		
ML	1.1588	123.4934	0.1406	0.2457
BE	1.1433	118.6451	0.1342	0.2959

Table 4.12: Estimates of Parameters for LND.

		Overall Crimes		
Methods	$\hat{\mu}$	$\hat{\sigma}^2$	**KS**	**P-values**
ML	9.8777	1.3341	0.2611	0.0009768*
		PUNJAB		
ML	9.4733	1.3525	0.22248	0.007902*
		SINDH		
ML	7.961592	1.225360	0.24142	0.00296*
		KPK		
ML	6.907075	2.181268	0.18176	0.05643
		BALOCHISTAN		
ML	5.673972	1.317147	0.14573	0.2016

		ISIAMBAD		
ML	5.231340	1.558098	0.1025	0.6217
		G.B		
ML	3.958365	1.431244	0.13761	0.2581
		AJK		
ML	4.907941	1.639256	0.19066	0.04242

Table 4.13: Estimates of Parameters for LGD.

		Overall Crimes		
Methods	$\hat{\alpha}$	$\hat{\beta}$	**KS**	**P-values**
ML	53.801	0.1836	0.2432	0.00269*
		PUNJAB		
ML	48.149	0.19675	0.20445	0.01863*
		SINDH		
ML	41.434	0.19215	0.22423	0.00724*
		KPK		
ML	9.8413	0.70185	0.15757	0.12273
		BALOCHISTAN		
ML	18.213	0.31153	0.13041	0.29136
		ISIAMBAD		
ML	11.064	0.47282	0.08898	0.75248
		G.B		
ML	7.5073	0.52727	0.10781	0.52179

The Table 4.1 gives the crime rates in Pakistan since 1997 to 2018. We can clearly see that the years 2010-12 had the highest crime rate in past 22 years.

The descriptive statistics of six crimes, each with 48 observations are presented in Table 4.2 and the descriptive statistics of \ sites wise crimes each with 54 observations can be seen in Table 4.3.

Table 4.2 indicates that other crimes which include dacoit, burglary and cattle theft are most committed with an average of 65232 and lowest committed crime is Murder with average 1519. Large variation is observed in the data that ranges from 3929294 to 104514081. The kurtosis in Table 4.2 ranges from 3.0269 to 6.3902, which is greater than 3, indicates that the distribution of crimes is Leptokurtic. Table 4.3 shows that the most crimes are committed in Punjab with average committed crimes as 44176, whereas, minimum number of crimes are committed in G.B with an average 168. Moreover, the average number of overall crimes which include the murder, attempt to murder, kidnapping, robbery, theft, and other crimes committed in Pakistan are 72070. It can also be observed that there is great variation in data ranges from 2193383 to 25916152144. The kurtosis ranges from 6.9914 to 7.3812, which is greater than 3, indicates that the distribution of crimes in different sites is Leptokurtic. Before fitting the different distributions, we check the basic assumptions of the distributions. All 48 observations in each crime are divided into two groups, each group contain 24 observations. The crimes in different sites contain 54 observations. The observations are divided into two groups with 27 observations in each. To assess the identicalness of crime data, Mann Whitney test is applied to crime data and results are used presented in Table 4.4 and 4.5. The corresponding P-values using Mann Whitney (U) test for each crime suggest the hypotheses of identical distributions at 5 percent level of significance are accepted. The corresponding P-values using Mann Whitney (U) test for the crimes in each site admire the hypothesis of identical distribution at the 5 percent level of significance are accepted. The values of Sindh are not identically distributed.

Tables 4.6, 4.7, 4.8, 4.9 showed the estimated parameters for FD, LLD, LND and LGD. Where $\hat{\alpha}$ is shape and $\hat{\beta}$ is scale parameters respectively. Table 4.6 represents the parameter estimates of crime data for FD distribution, by using ML and Bayesian methods. From the above table, we come to know that FD distribution is best fitted for Murder, Attempt to Murder, Kidnapping, Robbery, Other Theft and Others Crimes. The estimated parameters of LLD shape ($\hat{\alpha}$) and scale ($\hat{\beta}$) for murder, attempt to murder, kidnapping, robbery, other theft and other crimes are shown in Table 4.7. The estimated parameters of LLD by using ML and Bayesian methods. The results of P-values depict that LLD distribution is best fitted for all the variables except one variable, i.e. other crimes as its P-value are less than 0.05. Table 4.7 illustrates the estimated scale and shape parameter of the candidate distribution is estimated by using the ML method. The results show that all the variables are appropriate for the fitting of the LND but the variable other crimes are not fitted on the LND as its P-value is less than 0.05. The estimated parameters of LGD shape (α) and scale (β) for Murder, Attempt to Murder, Kidnapping, Robbery, Other Theft and Other Crimes are obtained in Table 4.10. The obtained estimated shape parameters are significantly different from each other and there is a significant difference in scale parameters. The p value is larger than 0.05 so MLE is a good method of estimation for murder, attempt to murder, kidnapping, robbery and other theft. The p value is smaller than 0.05 so MLE is not the best method of estimation for other crimes. Table 4.11 indicates the fitting of FD distribution on the site's data. Parameters of the given distribution are estimated by ML and Bayesian method. From the calculated results of the P-value at 5 percent level of significance we can see that FD is best fitted for all sites except overall crimes as its P-value is less than 0.05 which does not support the null hypothesis. Table 4.12 shows the fitting of LLD distribution for the crime rate in Pakistan. Parameters of the given distribution are estimated by using two methods, MLE and Bayesian. P-value at 5 percent level of significance is used to check the suitability of the distribution for overall crimes. It can be seen that LLD distribution is best fitted for all the sites, but for Islamabad it gives more significant results. Table 4.13 exhibits the estimates of the scale and shape parameters of LND along with their P-value of KS test at the 5 percent level of significance. It can be noted that LND distribution is good fitted only for KPK, Balochistan and Islamabad by using the only ML method. LGD distribution is fitted to the sites data and parameters are estimated by using ML methods. The results are presented in Table 4.12 for comparison purpose. The results of P-value depict that LGD distribution is best fitted for three sites, namely KPK, Balochistan and lslamabad.

Discussion

In the current study, data from different sites of Pakistan are collected for six crimes named as murder, attempt to murder, kidnapping, robbery, theft and other crimes. This study involves the frequency analysis of crime offenses in Pakistan using four parametric distributions which are FD, LLD, LND and LGD. We have put forward MLE and BE method for fitting our suggested distributions. The crime rates calculated in table 4.1 depicts the increase and decrease in crime rates during past 22 years. Pakistan had faced the highest crime rate in 2010,2011 and 2012 respectively. In the next years we see a decline in rate of crimes, but a surge can be clearly seen in year 2018. The descriptive statistics of six crimes, each with 48 observations are presented in Table 4.2. The results indicate that other crimes such as dacoit, burglary, assault and cattle theft are most committed with an average of 65232 and lowest committed crime is Murder with an average 1519. A large variation has been observed in crime data that ranges from 3929294 to 104514081. The kurtosis in Table 4.2 ranges from 3.0269 to 6.3902, which indicates that the distribution of crimes is Leptokurtic. The descriptive statistics of eight different sites each with 54 observations are presented in Table 4.3. The results showed that the most crimes are committed in Punjab with average committed crimes as 44176, whereas, minimum number of crimes are committed in G.B with average 168. Moreover, the average number of overall crimes which include the murder, attempt to murder, kidnapping, robbery, theft, and other crimes committed in Pakistan are 72070. A large variation has been observed in crime data that ranges from 2193383 to 25916152144. The kurtosis ranges from 6.9914 to 7.3812, indicates that the distribution of crimes in different sites is Leptokurtic. Before fitting the different distributions, we checked the basic assumptions of the distributions. All 48 observations in each crime are divided into two groups, each group contains 24 observations. The crimes in different sites contain 54 observations. The observations are divided into two groups with 27 observations in each. To assess the identicalness of crime data, Mann Whitney test has been applied to crime data and results are summarized in Table 4.3 and 4.4. The corresponding P-values using Mann Whitney (U) test for each crime suggests the hypotheses of identical distributions at 5% level of significance are accepted. This indicated that the distributions of all the crimes are identical. The corresponding P-values using Mann Whitney (U) test for the crimes in each site, except Sindh admire the hypothesis of identical distribution at the 5% level of significance are accepted. This shows that the distributions of crimes in all different sites are identical except Sindh. After the identicalness of the distribution, we checked the Skewness present in the observed data series by using test of Skewness. The values of coefficient of Skewness for different crimes are presented in Table 4.2 and Table 4.3 As all the values are greater than 0, indicating that observed data series are positively skewed. Further, Figures 4.1 to 4.5 also showed that the distribution of the crime data is skewed to right or positively skewed. Histograms showed that the distributions of the crimes are positively skewed. For the analysis of the crime data, we estimate the parameters by using ML and BE methods of the distribution, such as FD, LLD, LND and LGD. The P- value of KS test using both parameter estimation methods such as MLE and BE of data for different crimes in Pakistan indicates that FD is the best fit for all crimes. LLD, LND and LGD are a good fit for murder, attempt to murder, kidnapping, robbery and other theft but not a good fit for other crimes. Moreover, p-value of KS test using both parameter estimation methods such as MLE and BE of crime data for different site shows that FD is best fit to crime data in Punjab, Sindh, KPK, Balochistan, Islamabad, G.B and AJK but not a good fit for Overall crimes. LLD is best fit for all sites wise data. LND and LGD are best fit for KPK, Balochistan, Islamabad, G.B and AJK but not the best fit for Overall crimes, Punjab and Sindh. PP-Plots constructed for all crime data and also for crimes for all the sites also indicate the best fit of the selected distributions visually. The PDF plots of FD and LLD for murder, attempt to murder, kidnapping, robbery, other theft and other crimes are shown in figure 4.20 to 4.33. The graphs show that FD and LLD are best fit for crime wise and sites wise data.

Conclusion

Crime is a rising communal issue around the globe. The present study has been conducted in Pakistan to analyze crime counts by fitting the four different parametric distributions. Our proposed distributions for this study are FD, LLD, LNG and LGD. We have used MLE and BE methods. First, we have calculated the descriptive statistics like mean, variance, Skewness, kurtosis, minimum, maximum and C.V values of crime data with sample size n = 48. On the average the highest committed crime is other crimes as its frequency is 65232. Before fitting the distributions, we have verified the basic assumption of identicalness of distribution. To examine the identicalness of distribution we applied the Mann Whitney test. The results showed that distributions are identical. The R software is used for estimation of the parameters of our candidate distributions. While estimating the parameters of our lodged distributions using MLE and BE methods, the MLE of FD and LND are in closed form but BE are not in closed form. Therefore, an iterative method is used to obtain the estimates of the parameters. Moreover, the estimates of LLD and LGD are not in closed form. Therefore, an iterative method is used to obtain the estimates of parameters. To assess the performance of selected distributions of crime data, PDF plots for all the crime counts is constructed. Graphical conception plays an important role in any data analysis. For this purpose, histogram of Crime data with normal curves were also drawn. Histograms are illustrating the ranges of values of coefficient of Skewness. Before the selection of candidate distribution, we construct the PP-Plot to check the best fitting distribution on a data set. The PP - Plot is the graph of the empirical (CD) values plotted by using the theoretical CDF value. They are used to confirm whether hypothetical distributions are appropriate for crime data or not. For this purpose, FD, LLD, LND and LGD are selected. The PDF plots of FD, LLD, LND and LGD for murder, attempt to murder, kidnapping, robbery, other theft and other crimes are shown in figure 4.27 to 4.37. The graphs show that FD, LLD, LND and LGD are best fit for crime wise and sites wise data.

Recommendations

The current study focused on four distributions that are used for analysis of crime counts using MLE and Bayesian methods. Elongation of the study can be done by,

- Using other distributions
- Using other estimation methods such as PWM, ME, least square etc.
- Using data from more years
- Using time series analysis
- Conducting study in relation with areas

References

Wallace, W. C. (2017). *An exploratory study on the impact of organized crime on societies in Small Island Developing States: Evidence from five Caribbean countries.*

Akinrefon, A. A., Adeniyi, O. I., Adejumo, A. O., Olawale, A. O., & Ubong, B. A. (2016). Analysis of Criminal Cases in Adamawa State, Nigeria. *Global Journal of Social Sciences Studies*, *2*(3), 131-143.

Khan, N., Ahmed, J., Nawaz, M., & Zaman, K. (2015). The socio-economic determinants of crime in Pakistan: New evidence on an old debate. *Arab Economic and Business Journal*, *10*(2), 73-81.

Fitterer, J. L., & Nelson, T. A. (2015). A review of the statistical and quantitative methods used to study alcohol-attributable crime. *PloS one*, *10*(9), e0139344.

Khalid, S., Wang, J., Shakeel, M., & Nan, X. (2015, June). Spatio-temporal Analysis of the Street Crime hotspots in Faisalabad city of Pakistan. In *Geoinformatics, 2015 23rd International Conference on* (pp. 1-4). IEEE.