Parameter estimation of Koshi river basin rainfall data

Chinmay Sahu

Communications, Signal Processing, Networking Lab
Department of Electrical and Computer Engineering
Clarkson Univeristy,Potsdam,13699

April 12, 2019

Abstract

In this project, several parameters are estimated and studied from the recorded rainfall data of Koshi river basin. In particular, Climatological data over the Koshi river basin were recorded between 1975 and 2010. Precipitation data from 50 stations were collected. By analysis, it is observed that the recorded rainfall data follow a Laplace distribution. Tools are developed to estimate the maximum likelihood estimates of average rainfall and the spread for Laplace distributed data. The results from Laplace and Gaussian estimates are compared and verified using Cramer Rao lower bound(CRLB). The bias and variance in each case are evaluated and then the conclusions are drawn.

1 Introduction

The Koshi river basin is a sub-basin of the Ganges shared among China, Nepal, and India. Climatological data(Rainfall data) from 50 stations over the Koshi river basin were recorded between 1975 and 2010 [1]. For each day, the maximum rainfall across all 50 stations was recorded. From analysis, It is observed that the data recorded follow a Laplace distribution. First, the maximum likelihood estimators of location parameter(average) and the scale parameter(spread) for N iid observations are derived and then estimated for a randomly generated Laplace distribution data. The CRLB for both the estimators are evaluated. The sample mean and sample variance estimators derived for Gaussian distributed samples are used to estimate the mean and variance. Then bias and variance of each of the estimates are evaluated. Then the Gaussian estimates are compared with Laplace estimates and conclusions are drawn.

2 Results and Discussion

2.1 Parameter Estimation for Laplace distribution

A Laplace distribution follows a following function

$$f_X(x) = \frac{1}{2b}\exp(-\frac{|x-\mu|}{b}) \tag{1}$$

where μ is the location parameter and b is the scale parameter.

The ML estimators of the location parameter (average) and the scale parameter(spread) for N iid observations drawn from above Laplace distribution are given in equation 2 and equation 3 respectively.

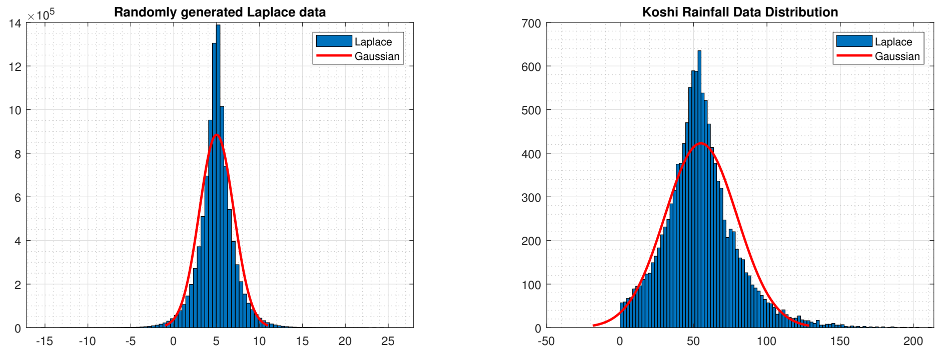
$$ML_{Estimator\hat{\mu}} = median(X_i) \tag{2}$$

$$ML_{Estimator\hat{b}} = \frac{1}{N}\sum_{i=1}^{n}|x-\mu| \tag{3}$$

The CRLB for location parameter μ and spread b are

$$CRLB_{\mu,b} = \frac{b^2}{N} \tag{4}$$

To verify the estimators, a large set randomly distributed Laplace data with $\mu = 5$ and sperad b=2 were generated in MATLAB. From the figure 1(a), It is observed that the dataset follow a Laplace distribution. The ML estimators are implemented in MATLAB and variance of each estimator is evaluated for different number of randomly generated samples through MonteCarlo simulations. The variance of μ reduced as the number of generated samples increased. For N=10, variance for μ is 0.2881, where CRLB for μ is 0.2. For N=100, variance for μ is 0.0229, where CRLB for μ is 0.02. For N=10^6, variance for μ is 1.9877×10^{-6}, where CRLB for μ is 2×10^{-6}. As the ML estimator for μ is asymptotically consistent, the μ achieves CRLB as N (number samples generated) increases.



(a) Distribution of randomly generated data (b) Distribution of Koshi Rainfall data

Figure 1: The above figure shows the distribution of randomly generated Laplace data and Koshi Rainfall data. It is visible from the sub-figures that both the data sets follow Laplace distribution.

The variance of $\hat{b}$ is

$$Var_{\hat{b}} = \frac{b^2}{N} \tag{5}$$

So, It is inferred that CRLB and Variance of $\hat{b}$ is same. From simulation, it is evident. For example, for N=10, the variance of $\hat{b}$ is 0.1999, where as the CRLB for $\hat{b}$ is 0.2. For N=100, the variance of $\hat{b}$ is 0.0199998, where as the CRLB for $\hat{b}$ is 0.02. For N=10^6, variance of $\hat{b}$ is 2.01×10^{-6}, where as the CRLB for $\hat{b}$ is 2×10^{-6}. It means that the ML estimator for $\hat{b}$ is efficient.

2.2 Model Mismatch

The sample mean and sample variance estimators for Gaussian distributed samples are used for Laplace distributed data. The sample mean for the randomly generated data is 5.0176. The sample variance for the randomly generated samples is 4.0271. The sample mean estimator is unbiased(Shown in attached scanned copy). The variance of the sample mean estimator is found to be(as shown in attached scanned copy)

$$Var_{\hat{\mu}} = \frac{var(x)}{N} \tag{6}$$

So the sample mean estimator is asymptotically consistent. The bias in the sample variance estimator is also found to be zero (shown in scanned copy). So the sample variance estimator is unbiased.The variance of sample variance estimator is derived to be

$$Var_{\hat{V}} = \frac{20b^4}{N} \tag{7}$$

So the sample variance estimator is also asymptotically consistent.

The variance of sample mean estimator reduced as the number of generated samples increased. For N=10, variance for sample mean is 0.39884, where CRLB value is 0.2. For N=100, variance of sample mean estimator is 0.04, where CRLB value is 0.02. For N=10^6, variance for sample mean estimator is 4.0214×10^{-6}, where CRLB for μ is 2×10^{-6}. As number of samples increases, the variance is decreasing. However, it is not able to achieve CRLB defined for Laplace distributed data.

Similarly, for sample variance estimator, as the N increases, the variance of the estimator reduces. However, it is not able to achieve the CRLB defined for Laplace distributed data. The ML estimators derived for Laplace distribution would be preferred over the sample mean and sample variance estimators as they achieve CRLB defined for them. So they result in consistently efficient estimators than the sample mean and sample variance estimators derived for Gaussian samples. As the provided rainfall data and randomly generated data follow a Laplace distribution, the ML estimators derived for Laplace samples would give better results than other estimators.

2.3 Rainfall Data

The estimators are used on Koshi rainfall data. The results are tabulated in Table 1.

	$\hat{\mu}$	$\hat{b}$
Laplace	53.3267	18.0328
Gaussian	55.0172	603.8253

Table 1: Estimators for Koshi Rainfall data

In case of Laplace, the ML estimator $\hat{\mu}$ is median of the data while for the Gaussian, it is sample mean. From Table 1, It is obvious that both mean and median are close to each other. While for $\hat{b}$, Laplace estimator is spread estimate and Guassion estimator estimates sample variance. Hence, There is visible difference in the results. The Laplace estimator estimates the spread while Gaussian estimator estimates the variance of Koshi rainfall data. As the Koshi rainfall data follows Laplace distribution, ML estimators derived for Laplace ditribution are preferred over Gaussian estimators. This would ensure an efficient and consistent estimation of location parameter $\hat{\mu}$ and spread $\hat{b}$ for given rainfall data.

3 Conclusions

The ML estimators for Laplace distributed data are derived. Then the CRLB for each estimators are derived. The estimators are implemented in MATLAB and results are compared with CRLB estimates. It is concluded that the ML estimator for location parameter $\hat{\mu}$ is asymptotically consistent. The ML estimator for spread $\hat{b}$ is efficient. The gaussion estimators (sample mean and sample variance estimators) are applied over the Laplace data. The Bias and variance of each of the estimators are derived. Then results are compared with CRLB for Laplace distributed data. Then both sets of estimators are used for Koshi rainfall data. Based on the results, ML estimators derived for Laplace distribution are preferred for provided that, as the provided data follows a Laplace distribution.

References

[1] Shrestha, A. B.; Bajracharya,S. R.; Sharma, A. R.; Duo,C.; and Kulkarni,A.,(1975-2010). Observed trends and changes in daily temperature and precipitation extremes over the Koshi river basin, International Journal of Climatology, 37(2), pp.1066-1083.

Appendix

Matlab Code Estimators

```
clc
clear

%%Laplace distribution Data generation
mu=5; %avergae
sd=2; %spread
N=1e+5;
y = Laprand_generator(1,N, mu, sd) ; % mu=5, sd=2
%LAPRND generate i.i.d. laplacian random number drawn from laplacian ...
    distribution
% with mean mu and standard deviation sigma.
% mu : mean
% sigma : standard deviation
% [m, n] : the dimension of y.
% Default mu = 0, sigma = 1.
% For more information, refer to
% http://en.wikipedia.org./wiki/Laplace_distribution

%% Laplace estimators
histfit(y,100);
grid on
grid minor
hold on
legend('Laplace','Gaussian');
title('Randomly generated Laplace data');

b=sd/sqrt(2);
ML_mu=median(y); % ML estimator of mu
ML_b= sum(abs(y-mu))/N; %ML estimator

%% CRLB Check for mu MonteCarlo
 for j=1:6
for i=1:100000
    m=10^j;
    y = Laprand_generator(1,m, mu, sd) ; % mu=5, sd=2
    Mu(i)=median(y);
    b(i)= sum(abs(y-mu))/m; %ML estimator
    SMean(i)=sum(y)/m;
    SVar(i)=(1/(m-1))*sum((y-SMean(i)).^2);
end
TCRLB_Mean(j)=var(SMean); % experimental
TCRLB_Var(j)=var(SVar);  % experimental
CRLB_Mu(j)=var(Mu); % experimental
CRLB_b(j)=var(b); %experimental
j
 end
```

```
CRLB_mu=2e-6;%(b*b)/N;

plot([10,100,1000,10000,100000,1000000],CRLB_Mu);
grid on
grid minor
hold on
plot(1000000,CRLB_mu,'r*');
legend('Variance of \Mu','Theoritical CRLB');
title('Theoritical CRLB vs Actual Variance of \mu ');


%% Gaussian mean and vaiance estimator
SampleMean=sum(y)/N;
SampleVar=(1/(N-1))*sum((y-SampleMean).^2);
Bias_mean=SampleMean-mu;
Var_mean=(1/N)*var(y);
Bias_Var=SampleVar-var(y);
CRLB_SampleMean=1/(N*var(y));
CRLB_SampleVar=(1/N)*(2*var(y)^2);

%% Koshi Basin data
x_k = xlsread('Koshi_Rainfall.csv');

indices=find(x_k<0);
x_k(indices)=[];
M=length(x_k);
histfit(x_k,100);
grid on
grid minor
hold on
legend('Laplace','Gaussian');
title('Koshi Rainfall Data Distribution');


LaplaceMean=median(x_k);
LaplaceSpread=(1/M)*sum(abs(x_k-LaplaceMean));
GaussianMean=(1/M)*sum(x_k);
GaussianVar=(1/(M-1))*(sum((x_k-GaussianMean).^2));
```