Research Article

An alternate method of evaluating Lagrange multipliers of MEP

SN

Zaheer Uddin[1] · Muhammad Bilal Khan[1] · Muhammad Hani Zaheer[1] · Waqar Ahmad[1] · Muhammad Ali Qureshi[1]

Abstract

We present an alternate method for evaluating Lagrange multipliers of probability distributions usually used in maximum entropy principle. The Lagrange multipliers are evaluated using least square method. Both the methods are used to calculate Lagrange multipliers in fitting wind speed data. Kolmogorov–Smirnov test, Chi squared goodness of fit test, and root mean square error indicate the usefulness of this alternate method.

Keywords Maximum entropy principle · Maximum entropy method · Wind data · Wind speed · Hansweert · Least square method · Least square fitting

1 Introduction

Maximum entropy principle has a variety of applications in modeling of stochastic data. These applications involve many branches of natural and social sciences. Pressé et al. [1] reviewed the origins and use of Maximum Entropy Principle in Statistical Physics and beyond in fields such as Biology. Karlin et al. [2] used MEP for constructing equilibria in lattice kinetic equations. Chen and Dai [3] investigated MEP of uncertainty distributions for uncertain variables. Županović et al. [4] took an alternative path to derive Kirchhoff's loop rule where they employed MEP to show that in network branches currents are distributed to achieve maximum entropy. Alves et al. [5] found a connection between the maximum entropy principle and the Higgs boson mass without introducing any extra assumptions into the Standard Model. Hanel etal. [6] generalized the notion of entropy to complex and non-Ergodic systems and showed that the MEP is a consistent method for such systems. Roux and Weare investigated the consistency of the MEP and restrained-ensemble simulations. They showed that the conditions of the restrained-ensemble simulations were compatible with the maximum entropy principle [7]. Aldana-Bobadilla and Kuri-Morales proposed a method of clustering based upon the MEP and showed their method

was more effective than supervised ones [8]. Pontzen and Governato used MEP to derive the phase-space distribution for a halo of dark matter. They used physically motivated constraints and found a good match with three simulations of dark matter [9]. Rahmati, Pourghasemi and Melesse applied random forest and maximum entropy models to map groundwater potential in the Mehran region of Iran and found MEP provided more successful predictions [10]. Ruggeri used non-linear MEP to model a polyatomic gas under changing pressure conditions. He found the framework to be in agreement with the phenomenological Extended Thermodynamics theory [11].

2 MEP and new method

For a probability density function $p(x)$, the entropy is defined as

$$H = -\int p(x)\ln p(x)dx \tag{1}$$

We need to maximize H subject to the condition that

$$E\{\phi_n(x)\} = \int \phi_n(x)p(x)dx = \mu_n, n = 0, \dots, N \tag{2}$$

here $\phi_n(x)$ re known functions with $\phi_0(x) = 1$ nd μ_n re moments about origin with $\mu_0 = 1$. The classical form of $p(x)$ is

$$p(x) = exp\left[-\sum_{n=0}^{N}\lambda_n\phi_n(x)\right] \tag{3}$$

here λ_n are called the Lagrangian parameters and make up a vector λ. For a given data set we need to find the value of N that best fits that data set, and hence we need to determine the distribution it follows. Following Algorithm is usually used to calculate the best fit distribution.

1. Define the range (x_{min}, x_{max}) and step size dx.
2. Use some standard functions for $\phi_n(x)$.
3. Start iterative procedure with some λ_o.
4. Calculate following two integrals.

$$G_n(\lambda) = \int \phi_n(x)exp\left[-\sum_{n=0}^{N}\lambda_n\phi_n(x)\right]dx$$
$$= \mu_n, n = 0, \dots, N \tag{4}$$

$$g_{nk} = g_{kn} = \frac{\partial G_n(\lambda)}{\partial \lambda_k} = -\int x^k x^n exp\left[-\sum_{m=0}^{N}\lambda_m x^m\right]dx$$
$$= -G_{n+k}(\lambda) \tag{5}$$

5. Solve the equation $\mathbf{G}\delta = \mathbf{v}$, where $v = [\mu_o - G_o(\lambda^o), \mu_1 - G_1(\lambda^o), \dots, \mu_N - G_N(\lambda^o)]^t$ to find δ
6. Calculate $\lambda = \lambda^0 + \delta$ The process is repeated with this new value as λ^0 ntil δ becomes negligible.

In this procedure, the first λ_o is basically an initial guess for the vector λ. Then the new values of λ are taken as the new $\lambda_o \cdot \delta$ are uncertainties in λ_o.

3 Least square method

In least square method, the sum of square of error (SSE) is made minimum. If y and ŷ are the observed and calculated values of a dependent variable respectively, the sum of square of error is given by

$$SSE = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{6}$$

The parameters of fitting are found by minimizing SSE.
To find the values of Lagrange multipliers, we have used least square fitting with $\phi_n(x) = x^n$. The classical form of $p(x)$ is reduced to a linear polynomial by taking logarithm of the equation.

$$y = -lnp(x) = \sum_{n=0}^{N}\lambda_n x^n \tag{7}$$

The values of Lagrange coefficients are found by minimizing sum of square of errors, i.e.

$$SSE = \sum_{i=0}^{N}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=0}^{N}\left(y_i - \sum_{n=0}^{N}\lambda_n x_i^n\right)^2 \tag{8}$$

These are $(N+1)$ linear equations in terms of Lagrange coefficients. They are solved to find the values of these coefficients. The process is repeated for various degrees of polynomials. The degree of the polynomial is selected for which Chi square, Kolmogorov–Smirnov test and Mean Square errors are minimal.

4 Wind speed data

Many different statistical distributions have been used to model wind speed data. The most widely used distribution is Weibull distribution [12, 13]. MEP more accurately models wind speed data [14–17]. Djafari developed a matlab program to model wind power density using maximum entropy distributions [18]. Wind speed data for this study has been downloaded from the site [19]. This site contains wind speed data of various cities of The Netherlands and it has been made available for free for educational and research purposes. As a sample we randomly picked the city, Hansweert and used hourly wind data for the year 2000. It has a longitude of 3.998 and latitude of 51.446. Hansweert is a village in southwest Netherlands with overall pleasant weather. The average temperature in winter is 2 °C, while in summer it is 18 °C. In Hansweert winters are windier.

5 Goodness of fit tests

Three different goodness of fit tests have been used to determine how well the datasets fit on the classical distribution of maximum entropy principle using both methods.

6 Kolmogorov–Smirnov test

Kolmogorov–Smirnov (KS) test is a goodness of fit test that checks if a given sample data follows a specific distribution. KS test measures the maximum distance between cumulative curves. It does not depend on cumulative distribution function (CDF). It is given by

$$KS = \max \left[F(x_i) - G(x_i) \right], i = 1, 2, 3, ...n \qquad (9)$$

where 'n' is the number of data points, $F(x_i)$ is the CDF of hypothetical or given distribution and $G(x_i)$ is the CDF of specific distribution.

To check the validity of the hypothesis that the given data follows a specific distribution, a test of hypothesis with given level of significance is conducted. If $KS < KS_\alpha$ then the two distributions are similar. At 99% level of significance, $KS_{99} = \frac{1.63}{\sqrt{n}}$. For n = 6 and $\alpha = 99\%$, the critical value is 0.618. The best feature of KS test is that it does not depend on sample size.

7 Chi square goodness of fit test

Chi square test is another goodness of fit test which also determines how well a given dataset follows a specific distribution. This test requires dataset of observed frequency (O_i) and corresponding expected frequency (E_i). The Chi square test statistic is given by

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \qquad (10)$$

If $\chi^2 < \chi^2_{\alpha,k-1}$ the two distributions are identical, α is the level of significance and k is the number of data points. For k = 6 and $\alpha = 95\%$, the critical value is 1.63.

8 Root mean square error (RMSE)

Root mean square error is another mean to determine the goodness of fit of a given dataset on a specific distribution. It is defined as the square root of mean of square error and is given by

$$\chi^2 = \sqrt{\sum_i \frac{(y_i - \bar{y}_i)^2}{n}} \qquad (11)$$

where y_i is the actual value and $\bar{y}_i$ is the fitted value calculated from given distribution, and n is the number of data points. The smaller the RMSE is, the better the goodness of fit.

9 Results and discussion

The probability of a real-valued random variable is always positive in a region 'R'. If the random variable follows the maximum entropy principle, the probability distribution is given by

$$p(x) = exp \left[- \sum_{n=0}^{N} \lambda_n x^n \right]. \qquad (12)$$

where λ_n are known as Lagrange coefficients and the values of these coefficients are determined using moments as constraints. In this study instead of using the moments as constraints, the least square method was used to fit a distribution on given wind speed data, that is, coefficients λ_n are determined by minimizing square of the difference of fitted and theoretical probabilities. The approach is different from maximum entropy method, even though the same classical probability distribution is fitted. A Python program has been developed for the fitting process, the code of which is given in "Appendix A" section.

The Chi square, Kolmogorov–Smirnov and Root Mean Square Error test statistics have been calculated for fitting process for datasets of each month from January to December. All three tests of goodness of fit (Chi Square, Kolmogorov–Smirnov, and Root Mean Square Error) yield very low values, favoring the hypothesis that the datasets follow the determined distributions. The KS and RMSE are found to be very close. In Fig. 1 a comparison between
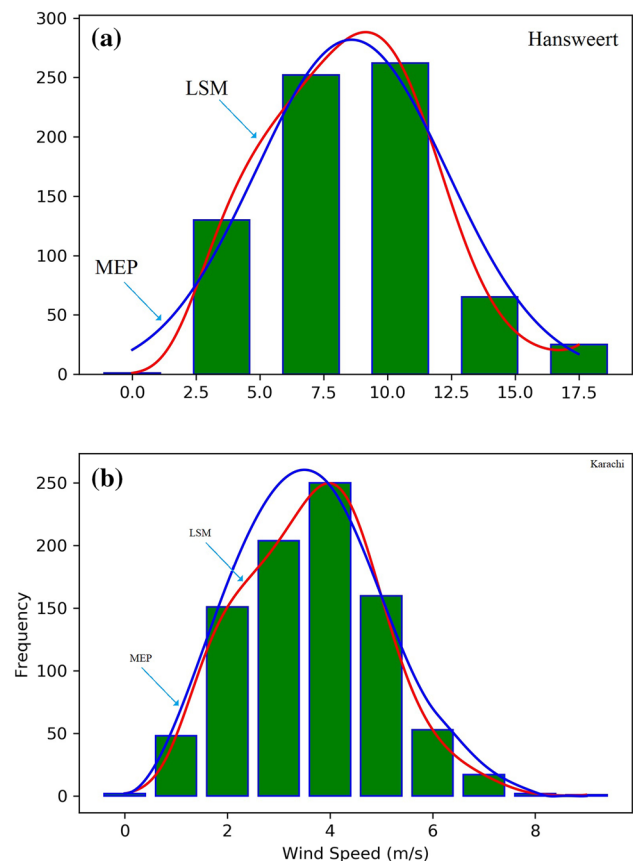

Fig. 1 Comparisons between Least Square fit and MEP fit for Hansweert (**a**) and Karachi (**b**). The histograms show actual data

Table 1 Goodness of fit statistics of MEP fit and least square fit for the month of January

Cities	Fits	Chi square	Kolmogorov–Smirnov	RMSE
Hansweert	MEP fit	0.058554	0.013204	0.029344
	Least square fit	3.05E−16	6.70E−08	6.86E−08
Karachi	MEP fit	2.48E−2	2.66E−3	3.62E−4
	Least square fit	5.63E−4	6.63E−4	1.263E−5

Table 2 Month-wise test statistics for goodness of fit for Least Square fit

Month	N	Chi square	Kolmogorov–Smirnov test	RMSE
January	5	3.05E−16	6.70E−08	6.86E−08
February	5	3.05E−16	6.70E−08	6.86E−08
March	5	7.88E−18	1.48E−08	1.26E−08
April	5	1.97E−17	1.85E−08	1.33E−08
May	5	1.52E−17	3.11E−08	1.49E−08
June	5	1.42E−16	3.92E−08	4.43E−08
July	5	3.35E−17	3.23E−08	2.54E−08
August	5	1.69E−17	2.04E−08	1.70E−08
September	5	8.65E−18	1.86E−08	1.17E−08
October	5	4.62E−16	1.35E−07	1.02E−07
November	6	1.44E−11	3.28E−05	1.64E−05
December	5	8.57E−18	3.63E−08	1.70E−08

MEP fit and least square fit of classical distribution of MEP is given for the wind speed data of month of January for Hansweert and Karachi, Pakistan. Both plots appear to be good fits for both cities. To compare the reliability of the fits, the values of test statistics for Chi square, KS and RMSE are given in Table 1. It is clear from the table that the values of these statistics are least in case of least square fit. The difference is not large for Karachi, but the superiority of the Least square fit is still apparent. The reason is that the main focus of MEP fitting is to determine exact values of the constraint parameters (moments) of the distribution. Hence if the statistical information of data is needed MEP is the better choice, and if fitting is the aim then least square fitting is the better choice.

In Table, values of all three test statistics for the wind speed data for January to December are given. On average the order of Chi square is 1E−16 but in November the order is 1E−11. The orders of KS and RMSE are almost same i.e. 1E−08, in November however the orders are 1E−5. The values show that the least square fitting of wind speed data on classical probability of MEP is excellent. The second column of Table 2 gives the degree of polynomial that fits in each month on the given dataset of wind speed. Mostly 6th degree polynomials fit well except for May & October where the degree of polynomial is 7. The Lagrange coefficients have been calculated and given in Table 3. These coefficients are then used to draw plots of classical probability distributions, which are overlapped on histograms generated from given wind speed data. In Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13, the plots show the least square fitted curves on

Table 3 Values of Lagrange coefficients obtained by Least Square Method

	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
Jan	0.2586	−3.7199	1.1559	−0.1965	1.7840E−02	−7.9752E−04	1.3813E−05	
Feb	−0.0962	−2.6092	0.7384	−0.1229	1.0853E−02	−4.6586E−04	7.7139E−06	
Mar	−2.2530	−1.4277	0.2821	−0.0390	3.7522E−03	−1.9167E−04	3.8228E−06	
Apr	−1.4752	−1.4661	0.0975	0.0120	−1.5274E−03	4.6887E−05		
May	−8.5501	8.9182	−4.8420	1.0956	−1.2571E−01	7.7100E−03	−2.4044E−04	2.9919E−06
Jun	0.6504	−4.0858	1.2287	−0.2112	2.1028E−02	−1.0907E−03	2.2989E+00	
Jul	−1.2563	−2.3945	0.5998	−0.0938	9.9086E−03	−5.8817E−04	1.4429E−05	
Aug	0.1333	−5.0964	1.9391	−0.3807	3.9318E−02	−1.9633E−03	3.7214E−05	
Sep	−0.5050	−2.7926	0.6578	−0.0698	1.2434E−03	3.0950E−04	−1.5337E−05	
Oct	2.4246	−3.9655	0.9971	−0.1661	1.7848E−02	−1.1256E−03	3.7408E−05	−5.0212E−07
Nov	38.0074	−31.1175	9.8836	−1.6705	1.5398E−01	−7.2824E−03	1.3832E−04	
Dec	3.1023	−3.7479	0.7224	−0.0822	5.5612E−03	−1.9647E−04	2.7869E−06	

Fig. 2 Histogram and LS fit for January

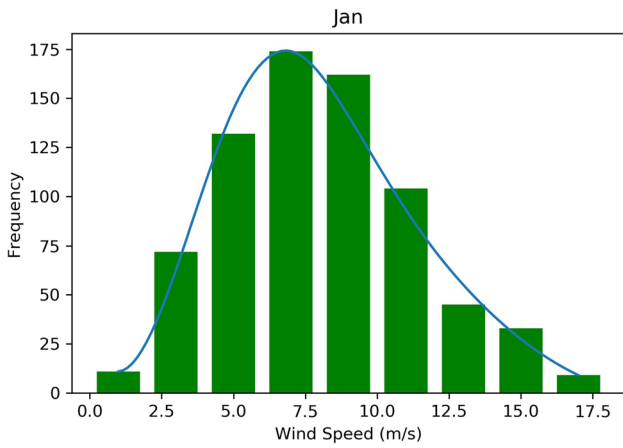**Fig. 3** Histogram and LS fit for February

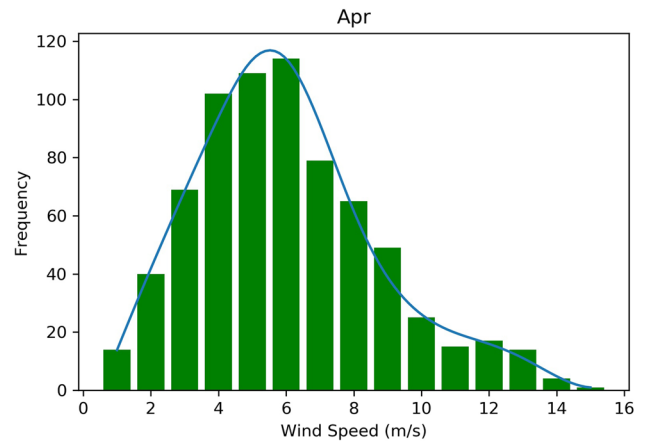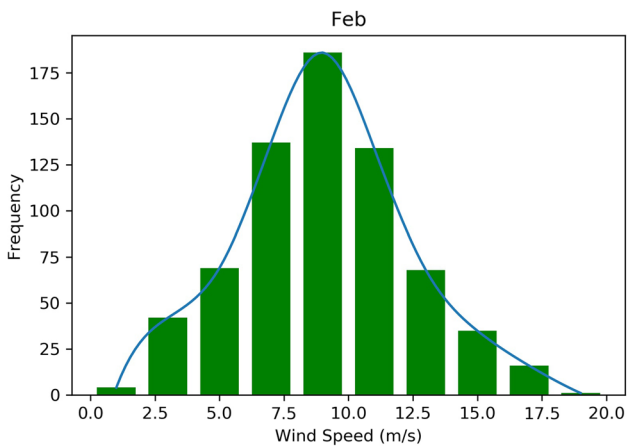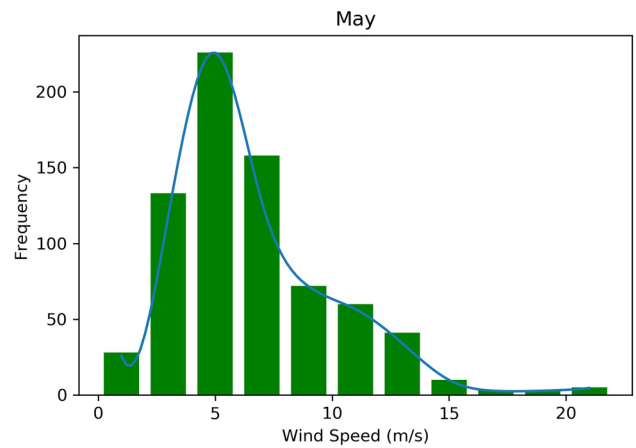**Fig. 4** Histogram and LS fit for March

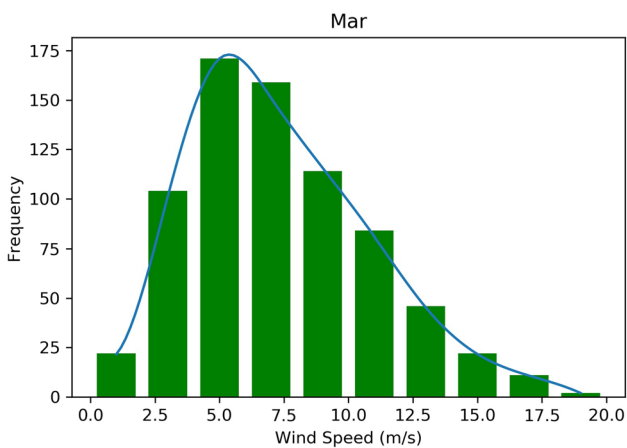**Fig. 5** Histogram and LS fit for April

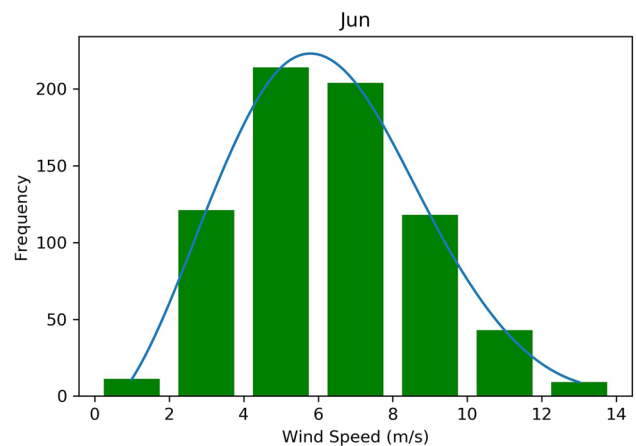**Fig. 6** Histogram and LS fit for May

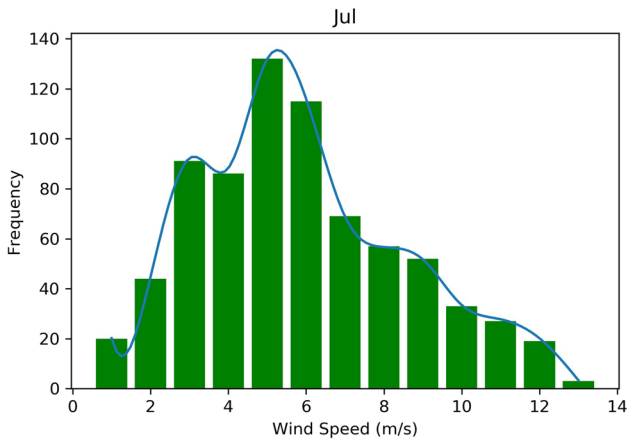**Fig. 7** Histogram and LS fit for June

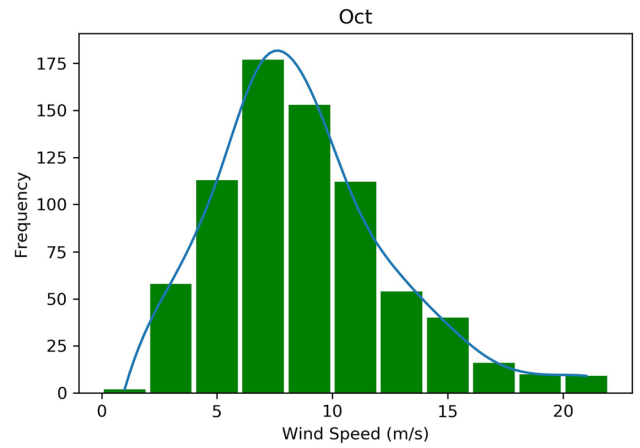**Fig. 8** Histogram and LS fit for July

Fig. 11 Histogram and LS fit for October
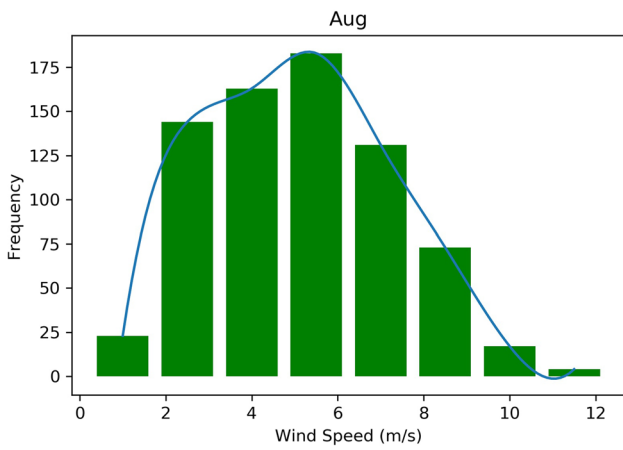
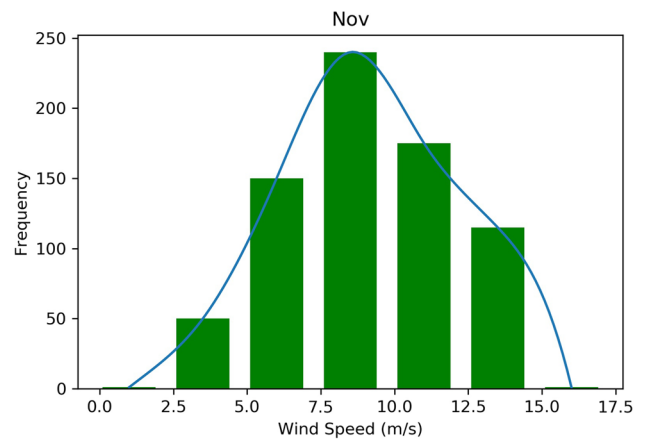**Fig. 9** Histogram and LS fit for August

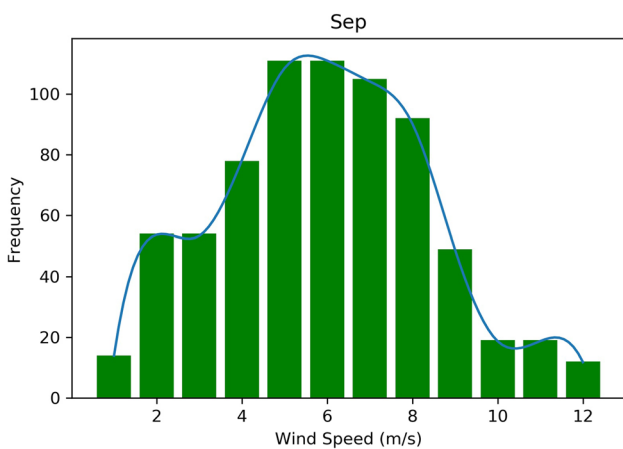**Fig. 12** Histogram and LS fit for November

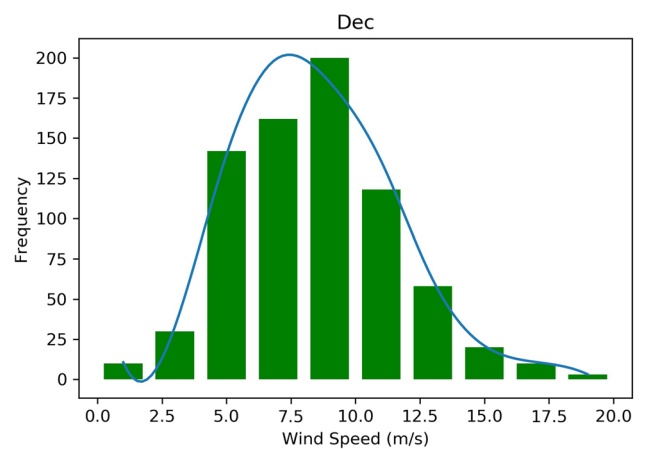

Fig. 10 Histogram and LS fit for September

Fig. 13 Histogram and LS fit for December

the given dataset. The plots show that the fitting of the dataset of wind speed on the probability distribution is remarkable.

10 Conclusion

In this study, we evaluated Lagrange multipliers of classical probability distribution of MEP using Least Square Method on wind speed data. The hourly wind speed data of Hansweert (The Netherlands) from January to December (2015) is used in this study. The basic difference between MEP and the method used is that whereas MEP fits the distribution using moments as constraints, the alternative method minimizes the square of differences between theoretical and given probabilities. It is concluded that

(i)   MEP should be used if statistical information of the distribution is needed, and LSM is preferable when fitting of dataset is required.

(ii) The results show the evaluation of Lagrange multipliers through LSM is better than that obtained by MEP (see Table 1).

A Python program has been developed that fits various degree polynomials as classical probability distributions and displays the degree for which the errors are least. It also displays the calculated values of Lagrange multipliers. For error calculation Chi square, Kolmogorov–Smirnov and Root Mean Square Error test statistics were used. The alternative method consists of using least square method to find the Lagrange multipliers of maximum entropy principle and has proven to provide a better fit of the dataset, as evidenced by lower values of tests of significance statistics.

Compliance with ethical standards

Appendix

```
" Least Square fitting of MEP probability distribution"

import matplotlib.pyplot as pyplot

import numpy as np

from numpy.linalg import inv

import xlrd

from scipy.interpolate import spline

""" Open file for input data"""

file_location="C:/*.xlsx"

workbook=xlrd.open_workbook(file_location)

sheet=workbook.sheet_by_index(0)

x=[]

y=[]

p=[]

for row in range(sheet.nrows):

    x.append(sheet.cell_value(row,0))

    y.append(sheet.cell_value(row,1))

RMSE=[]

for n in range(2,12):

    if (n==11):

        n=int(RMSE.index(np.min(RMSE)))+2

        print('Degree = ',n,'RMSE =',RMSE[n-2])

    for i in range(0,len(x)):

        if (y[i]==0):

            y[i]=1

        p.append(-1*(np.log(y[i])))

    y1=p

    """ Least square fitting"""

    y2=[]

    coeff=[]

    xy=[]

    for k in range(0,n+1):

        temp1=0

        row=[]

        for i in range(0,n+1):
```

```
            temp=0
            for j in range(0,len(x)):
                temp=temp+(x[j]**i)*(x[j]**k)
            row.append(temp)
        coeff.append(row)
    for k in range(0,n+1):
        temp1=0
        for i in range(0,len(x)):
            temp1=temp1+(x[i]**k)*y1[i]
        xy.append(temp1)
    ainv=inv(coeff)
    a=np.matmul(ainv,xy)
    """ Calculate expected frequency """
    for i in range(0, len(x)):
        temp=0
        for j in range(0,n+1):
            temp=temp+a[j]*x[i]**j
        y2.append(np.exp(-1*temp))
    yfit=[]
    for i in range(0,len(x)):
        yfit.append(y2[i]*np.max(y)/np.max(y2))
    #plt.plot(x,y)
    "Calculates RMSE"
    sum=0
    for i in range(len(x)):
        sum=sum+(y[i]-yfit[i])**2
        #print(y[i],yfit[i])
    RMSE.append(sum/len(x))
""" Plots histogram from input dta and polygon from fitted values"""
x2=np.linspace(x[0],x[len(x)-1],100)
dx=(x[len(x)-1]+1)/1000
ynew=[]
xnew=[]
xtemp=-dx
```

```
for i in range(0,1000):

    xtemp=xtemp+dx

    sum=0

    for j in range(0,n+1):

        sum=sum+a[j]*xtemp**j

    ynew.append(np.exp(-sum))

    xnew.append(xtemp)

y3=spline(x,yfit,x2)

pyplot.plot(x2,y3)

pyplot.bar(x,y,color="green",edgecolor='red',width=1.45)
```

References

1. Pressé S, Ghosh K, Lee J, Dill KA (2013) Principles of maximum entropy and maximum caliber in statistical physics. Rev Modern Phys 3(85):1115
2. Karlin IV, Gorban AN, Succi S, Boffi V (1998) Maximum entropy principle for lattice kinetic equations. Phys Rev Lett 81(1):6
3. Chen X, Dai W Maximum entropy principle for uncertain variables. Int J Fuzzy Syst 13(3):232–6 (2011)
4. Županović P, Juretić D, Botrić S. Kirchhoff's loop law and the maximum entropy production principle. Phys Rev E 70(5):056108 (2004)
5. Alves A, Dias AG, da Silva R (2015) Maximum entropy principle and the Higgs boson mass. Phys A Stat Mech Appl 7(1):420
6. Hanel R, Thurner S, Gell-Mann M (2014) How multiplicity determines entropy and the derivation of the maximum entropy principle for complex systems. Proc Natl Acad Sci 111:6905–6910
7. Roux B (2013) Weare J On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. J Chem Phys 138(8):02B616
8. Aldana-Bobadilla E, Kuri-Morales A (2015) A clustering method based on the maximum entropy principle. Entropy 17(1):151–180
9. Pontzen A, Governato F (2013) Conserved actions, maximum entropy and dark matter haloes. Mon Notices R Astronom Soc 1(430):121–133
10. Rahmati O, Pourghasemi HR, Melesse AM (2016) Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. CATENA 137:360–372
11. Ruggeri T (2015) Non-linear maximum entropy principle for a polyatomic gas subject to the dynamic pressure. arXiv preprint arXiv:1504.05857
12. Khan JK, Shoaib M, Uddin Z, Siddiqui IA, Aijaz A, Siddiqui AA, Hussain E (2015) Comparison of wind energy potential for coastal locations: Pasni and Gwadar. J Basic Appl Sci 11:211
13. Khan JK, Ahmed F, Uddin Z, Iqbal ST, Jilani SU, Siddiqui AA, Aijaz A (2015) Determination of Weibull parameter by four numerical methods and prediction of wind speed in Jiwani (Balochistan). J Basic Appl Sci 11:62
14. Zhang H, Yu YJ, Liu ZY (2014) Study on the maximum entropy principle applied to the annual wind speed probability distribution: a case study for observations of intertidal zone anemometer towers of Rudong in East China Sea. Appl Energy 114:931–938
15. Sobczyk K, Trcebicki J (1999) Approximate probability distributions for stochastic systems: maximum entropy method. Comput Methods Appl Mech Eng 168(1–4):91–111
16. Chellali F, Khellaf A, Belouchrani A, Khanniche R (2012) A comparison between wind speed distributions derived from the maximum entropy principle and Weibull distribution. Case of study; six regions of Algeria. Renew Sustain Energy Rev 16(1):379–385
17. Gurley KR, Tognarelli MA, Kareem A (1997) Analysis and simulation tools for wind engineering. Probab Eng Mech 12(1):9–31
18. Mohammad-Djafari A (1992) A Matlab program to calculate the maximum entropy distributions. In: Maximum entropy and Bayesian methods (pp. 221–233). Springer, Dordrecht
19. http://projects.knmi.nl/klimatologie/onderzoeksgegevens/potentiele_wind. Accessed July 2018