STATISTICAL ANALYSIS OF VARIABLES IN A POWER SYSTEM

[1]Geethu Joseph, [2]P.C.Thomas
[1]PG student, [2]Professor
Amal Jyothi College of Engineering, Kanjirapally, Kottayam, Kerala
Email: [1]geethubjose@gmail.com, [2]pcthomas@amaljyothi.ac.in

Abstract— The variables in a power system of Amaljyothi, an educational institute, is statistically analysed to understand the correlation between them. The power system consist of grid supply, 100kW solar plant. The random variables identified are monthly energy demand, monthly energy consumption, net payable charges to KSEB and solar energy output. Also suitable probability distributions are fitted to these variables to identify the distribution pattern. There are different methods of fitting of probability distribution and appropriate fitting is tested using several goodness of fit tests. Easy fit software is used for fitting of probability distribution.

Index Terms— Probability distribution, random variable, distribution fitting , goodness of fit.

I. INTRODUCTION

The integration of more and more renewable energy sources into power systems brings in new challenges for planning and operation. Hence an accurate estimation of various components of the power system is necessary. The estimation of proper distribution parameters will help in predicting the potential of renewable energy sources in power system and also for accommodation of several random variables associated with future addition of more renew-able energy sources into the power system. Due to the random nature of weather conditions, it is appropriate to choose probabilistic methods rather than deterministic methods in analysis of the power system [1].

The ability to accurately assess renewable energy resources is an essential prerequisite to integrating renewable energy technologies into the energy supply portfolio of any community. In order to analyze the random variables of a renewable energy power system, the identification of proper distribution that fits these variables are necessary. Distribution fitting is the term that is relevant in this context. Distribution fitting is the procedure of selecting a statistical distribution that best fits to a data set generated by some random process. Apart from fitting the distribution to a data, the parameters of the distribution should be estimated. This can be done using several methods , the accuracy of which can be compared using various goodness of fit tests.The different probability distribution analysed are Normal, Uniform, Gamma, Beta ,Rayleigh ,Burr ,Exponential and Weibull .
Goodness of fit test used includes Anderson-Darling test, Chi-square test, Kolmogorov-Smirnov test. The parameters of the probability distribution are identified through Maximum likelihood Method.

II. PROBABILITY DISTRIBUTION , FITTING AND PARAMETER ESTIMATION

The probability distributions are better represented by probability density functions (PDF) which represents the relative likelihood for a random variable to take on a given value.

Some of the probability distributions used in literatures include Normal distribution with parameters (μ,σ), Uniform distribution with parameters (a, b),Gamma distribution with parameters (α,β,γ) ,Beta distribution with parameters(α,β) ,Rayleigh distribution with parameter (σ, γ),Exponential distribution with parameter λ ,Burr distribution with parameters(k,c),Weibull distribution with parameters(α,β,γ).

Distribution fitting is a procedure for selecting a statistical distribution that fits to a data set. The goodness of fit tests includes Chi-square test, Kolmogrov-smirnov test and Anderson –Darling test.

1) Chi square test

In order that we apply chi square test ,it is necessary that the observed as well as expected frequencies must be grouped in the same way and the theoretical distribution must be adjusted to give the same total frequency as we find in case of observed distribution [2].The test statistics, χ^2 is calculated as

$$\chi^2 = \sum \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

(1)

Where O_{ij} is the observed frequency, E_{ij} is teexpected frequency .The fit is considered a good one if the calculated value of χ^2 is less than the critical value of test at a particular level of significance.

2) Kolmogrov-Smirnov test

It involves a comparison between the theoretical and experimental cumulative distribution functions [3].The test statistics, D, is calculated as

$$D=||_{1\le i\le N}^{max}(F(Y_i) - \frac{n(i)}{N})$$

(2)

Where F(Y_i) forms the experimental cumulative distribution

n(i) is the number of points less than Y_i and N forms the total no .of observations .The fit is considered good if the test statistic 'D' is less than the critical value of the test at a certain level of significance.

3) Anderson Darling test

The test is defined by

$$A^2 = -N - S$$

(3)

Where N is the no. of observations and

$$S = \sum_{i=1}^{N} \frac{2i-1}{N}[\ln F(Y_i) + \ln(1 - F(Y_{N+1-i})]$$

(4)

The fit is considered to be good if the calculated value of test statistics, 'A' is less than critical value of test at a particular level of significance [4].

There are several methods for identification of distribution

Parameters such as Graphical method, Method of moments, Maximum likelihood method etc. The method used in the paper is Maximum likelihood Method because it has the advantage that as the no. of observations increases, the estimate converges to true values. Let $v_{1,}v_{2,}v_3 \ldots v_n$ forms a random sample drawn from a probability density function where θ is an unknown parameter .The likelihood function is defined as

$$L = \prod_{n}^{i=1} f_{vi}\,(v_i,\theta)$$

(5)

The maximum likelihood estimator of $\theta, \bar{\theta}$ is the value of θ that maximises L [5]. The method involves estimation of parameters through numerical iterations.

III. DATA COLLECTED

The data are recorded for three time zones. The first zone forms the normal time from 6 a.m to 6 p.m. The second zone forms the peak time from 6 p.m to 10 p.m and third zone forms the off peak time from 10 p.m to 6 p.m. The essential variables collected from KSEB(Kerala State Electricity Board) bill includes Demand in kVa for three time zones, Energy consumption in three zones in kWh, Net payable charges to KSEB and Total energy consumption in kWh. The details collected for year 2013 is shown in table 1and table 2.Similarly the data were collected from 2010 to 2015.

Table 1:Grid details for 2013

Month	Demand in zone 1 (kVa)	Demand in	Demand in

		Zone 2 (kVa)	Zone 3(kVa)
Jan	254	210	189
Feb	232	188	175
March	270	197	193
Apr	285	216	201
May	212	208	176
June	246	210	188
July	205	205	184
Aug	206	215	179
Sep	284	227	201
Oct	299	238	210
Nov	211	199	166
Dec	271	249	209

Table 2: Grid details for 2013

Month	Energy In zone 1(kWh)	Energy in zone 2(kWh)	Energy In zone 3(kWh)	Total KSEB Energy (kWh)	Net Payable Charges (Rs)
Jan	41955	7851	24228	804371	74034
Feb	35073	6198	23058	641479	64329
Mar	37512	7653	24129	753031	69294
Apr	38016	12624	28251	903518	78891
May	31365	13410	26127	766276	70902
June	35256	12733	23812	746849	71801
Jul	28083	12555	18342	544641	58980
Aug	32190	16158	23901	654735	72249
Sep	33135	14160	20592	661639	67887
Oct	43521	18915	25869	834036	88305
Nov	19833	7869	11418	390927	39120
Dec	39762	18081	26454	782466	84297

The power system consists of a 100 kW solar power plant (no battery storage) that started functioning on April 2013. The monthly readings of energy production from May 2014 to March 2015 are shown in Table 3.

Table 3: Solar energy generation details

Month(2014-2015)	Solar energy output (kWh)
May	5781
June	7032
July	6405
August	6378
September	8220
October	7190
November	6378
December	8251
January	9827
February	8839
March	10044

IV. CORRELATION BETWEEN GRID VARIABLES

The correlation was done to find a relation between variables in a power system .The variables included are Demand in kVa for the three time zones, Energy consumption in kWh for 3 time zones. Total energy consumption in kWh , Net payable charges to KSEB in Rs. and Solar energy output in kWh. The month wise data from 2010 to 2015 was used. The correlation is found using the Pearson's correlation coefficient 'r' given by

$$r=\frac{\sum(X_i-X_{avg})(Y_i-Y_{avg})}{\sqrt{\sum(X_i-X_{avg})^2\ \sum(Y_i-Y_{avg})^2}} \qquad (6)$$

where X_i , Y_i denotes the variables and X_{avg}, Y_{avg} denotes their average values. The correlation values are shown in Table 3.D1,D2,D3 denotes demand in 3 zones,E1,E2,E3 denotes energy in 3 zones, SP denotes solar power, TE denotes total energy consumption from KSEB, Cost denotes the net payable charges to KSEB.

Table 4: Correlation between the variables

	D1	D2	D3	E1	E2	E3	SP	TE	Cost
D1	1	.94	.96	.76	.56	.91	.23	.87	.85
D2	.94	1	.93	.78	.64	.93	.18	.90	.80
D3	.96	.93	1	.80	.60	.94	.17	.93	.81
E1	.76	.78	.80	1	.27	.86	.25	.96	.86
E2	.56	.64	.60	.27	1	.57	.43	.67	.64
E3	.91	.93	.94	.85	.57	1	.11	.91	.49

SP	.23	.18	.17	.25	.43	.11	1	.05	.44
TE	.87	.90	.93	.96	.67	.91	.05	1	.76
Cost	.85	.80	.81	.86	.64	.49	.44	.76	1

The following conclusions can be made

1. Among the demand for 3 time zones, SP has the highest correlation with demand of zone 1.This is because SP energy is utilized only during time zone 1 and there is no battery backup for a later use at night.
2. Considering the energy consumption for 3 zones ,
   TE is highly correlated with E1 which is attributed to the high energy consumption during time zone 1 due to the construction works, classes and lab sessions.
3. The tariff rates of energy are higher for zone 2 followed by zone 1 and zone 3.However energy consumption is highest for zone 1.So cost has a high correlation with E1 followed by E2 and then E3.
4. Considering the years 2011-2012-2013, tarrif rates increased yearly. When the KSEB bill was analysed, it was found that for Jan 2011 to June 2012, tariff rates for demand charges were 175, 87.5 and 93.3 for 3 time zones respectively. From July 2012 to April 2013, tariff rates for demand were 200,100 and 106.66 for 3 time zones respectively. From May 2013 onwards tariff rates are considered as Rs .400 per kVa .Demand charges are considered high for zone 1 and lowest for zone 2.Also from the correlation results, the D1 has a high correlation with net payable charges and for D2 has the lowest correlation.

V. PROBABILITY DISTRIBUTION FITTING RESULTS

The variables identified are fitted using the Easy Fit software. The fitting is tested using chi-square test ,kolmogrov-smirnov test and Anderson darling test and the distribution that is well accepted by the three tests is chosen. Also the parameter of the distribution is calculated using Maximum likelihood method. The probability density function associated with each variable is thus obtained.

1) Demand for Zone 1

The distribution representing demand in zone 1 is 2 parameter Rayleigh distribution with parameters σ=6.6798 and $\gamma = 136.08$

Table 5:Fitting results for demand in zone 1

Test	Test statistics	Critical value at 0.05 level of significance
Kolmogrov smirnov	0.0808	0.18174
Anderson Darling	0.57659	2.5018
Chi-square	2.9145	9.4877

Probability density function associated with demand in
zone 1 is

$$f(x)=\frac{x-136.08}{4461.97}\exp(-0.5*\frac{(x-136.08)^2}{4461.97}) \quad (7)$$

2) Demand for zone 2

The distribution representing demand for zone 2 is the 2 parameter Rayleigh distribution with σ=55.49 and $\gamma = 115.95$

Table 6: Fitting results for demand in zone 2

Test	Test statistics	Critical value at 0.05 level of significance
Kolmogrov smirnov	0.08948	0.18174
Anderson Darling	0.6516	2.5018
Chi-square	3.7974	11.07

Probability density function associated with demand in zone 2 is

$$f(x)=\frac{x-115.95}{3079.14}\exp(-0.5*\frac{(x-115.95)^2}{3079.14}) \quad (8)$$

3) Demand for zone 3

The distribution representing demand for zone 3 is the 4 parameter beta distribution with parameters

$$\propto_1 = 1.351, \propto_2 = 1.9022, a = 103.99, b = 248.34$$

Table 7:Fitting results for demand in zone 3

Test	Test	Critical

	statistics	value at 0.05 level of significance
Kolmogrov smirnov	0.0653	0.8174
Anderson Darling	0.3349	2.5018
Chi-square	0.60029	11.07

Probability density function associated with demand in zone 3 is

$$f(x)=2.98*\frac{(x-103.99)^{0.351}(248.34-x)^{0.9022}}{3254.6} \quad (9)$$

4) Energy consumption for zone 1

The distribution representing energy for zone 1 is the 2 parameter gamma distribution with parameters $\propto = 26.515, \beta = 1216.4$

Probability density function associated with energy of zone 1 is

$$f(x)=\frac{x^{25.515}\exp(\frac{-x}{1216.4})}{1216.4^{26.515}*8.26*10^{25}} \quad (10)$$

Table8:Fitting results for energy in zone 1

Test	Test statistics	Critical value at 0.05 level of significance
Kolmogrov smirnov	0.03282	0.8174
Anderson Darling	0.13692	2.5018
Chi-square	0.93539	11.07

5) Energy consumption for zone 2

The distribution representing energy for zone 2 is the gamma distribution with $\alpha = 6.0085, \beta = 1277.5, \gamma = 4131.9$

Table 9:Fitting results for energy in zone 2

Test	Test statistics	Critical value at 0.05 level of significance
Kolmogrov smirnov	0.06954	0.8174
Anderson Darling	0.31018	2.5018
Chi-square	5.4677	11.07

The probability density function associated with energy for zone 2 is

$$f(x)=\frac{(x-4131.9)^{5.0085}}{1277.5^{6.0085}*121.753} exp\frac{-(x-4131.9)}{1277.5} \quad (11)$$

6) Energy consumption for zone 3

The distribution representing energy for zone 3 is 3 parameter Weibull distribution with parameters $\alpha = 2.5649, \beta = 14062, \gamma = 6806.2$

Table 10:Fitting results for energy in zone 3

Test	Test statistics	Critical value at 0.05 level of significance
Kolmogrov smirnov	0.05835	0.8174
Anderson Darling	0.19487	2.5018
Chi-square	0.28489	11.07

The probability density function associated with energy for zone 3 is

$$f(x)=\frac{2.5649}{14062}\frac{(x-6806.2)^{1.5649}}{14062} exp\frac{-(x-6806.2)^{2.5649}}{14062} \quad (12)$$

6) Total energy consumption from grid

The distribution representing total energy is the 2 parameter gamma distribution with parameter $\propto = 6.0952, \beta = 82685$.The probability density function associated with the total energy consumption from grid is

$$f(x)=\frac{x^{5.0952}\exp(\frac{-x}{82685})}{82685^{6.0952}*141.278} \quad (13)$$

Table 11.Fitting results for Total energy

Test	Test statistics	Critical value at 0.05 level of significance
Kolmogrov smirnov	0.0852	0.8174
Anderson Darling	0.9641	2.5018
Chi-square	4.0701	11.07

7) Net payable charges to KSEB

The distribution representing net payable charges is the 3 parameter gamma distribution with $\alpha = 31.584, \beta = 2348.8, \gamma = 10837$

Table 12:Fitting results for net payable charges

Test	Test statistics	Critical value at 0.05 level of significance
Kolmogrov smirnov	0.05387	0.8174
Anderson Darling	0.16942	2.5018
Chi-square	2.5856	11.07

The probability density function associated with net payable charges to KSEB is

$$f(x)= \frac{(x-10837)^{30.584}}{2348^{31.584}*1.962*10^{33}} exp \frac{-(x-10837)}{2348.8} \quad (13)$$

8) Solar energy output

The distribution representing solar energy output is the 3 parameter gamma distribution with $\alpha = 1.75, \beta = 1190.4, \gamma = 5612.1$

Table 13: Fitting results for solar energy output

Test	Test statistics	Critical value at 0.05 level of significance
Kolmogrov smirnov	.0843	0.391
Anderson Darling	0.24776	2.5018
Chi-square	0.27191	3.8415

The probability density function associated with energy for zone 2 is

$$f(x)= \frac{(x-5612.1)^{0.75}}{1190^{1.75}*0.92} exp \frac{-(x-5612.1)}{1190.4} \quad (14)$$

VI.CONCLUSION

Statistical analysis of the variables in a power system was done to find out the correlation between the variables. The fitting of probability distribution helped in understanding the distribution pattern of each variable. From the fitting various results were obtained

a) Demand for zone 1 follows Rayleigh distribution.
b) Demand for zone 2 follows Rayleigh distribution .
c) Demand for zone 3 follows Beta distribution .
d) Energy consumption for zone 1 follows Gamma distribution.
e) Energy consumption for zone 2 follows Gamma distribution.
f) Energy consumption for zone 3 follows Weibull distribution.
g) Total energy consumption from grid follows Gamma distribution.
h) Net payable charges to KSEB follow Gamma distribution.
i) Solar energy output follows Gamma distribution.

REFERENCES

[1] S. Negi and L. Mathew, "Hybrid Renewable EnergySystems :A Review", International Journal of Electronic and Electrical Engineering, vol.7, no.5, pp.535-542, 2014.

[2] D.Indhumathy , C.V.Seshaiah and K.Sukkiramathi, "Estimation of Weibull Parameters for Wind speed calculation at Kanyakumari in India", International Journal of Innovative Research in Science, Engineering and Technology , vol.3, pp.8340-8345, 2014.

[3] H,Garba, A.Ismail and U.Tsoho," Fitting Probability Distribution to discharge variability of Kaduna River",International Journal of Modern Engineering Research,vol.3,pp.2848-2852,2013.

[4] H. Cherif and J. Belhadj,"Energy output estimation of hybrid Wind-Photovoltaic power system using statistical distributions", Journal of Electrical systems , vol.10, no.2,pp.117-132, 2014.

[5] A. Chauhanl, R. P. Saini,"Statistical Analysis of Wind Speed Data Using Weibull Distribution Parameters",IEEE International Conference on Non Conventional Energy,vol.5,pp.160-163, 2014