Validation of Beta Distribution for Spectrum Usage using Kolmogorov-Smirnov Test

Kishor Patil Sinhgad Academy of Engineering Savitribai Phule Pune University Pune

ABSTRACT

The arrival and departure of licensed users in the licensed band is a random process. So the channel availability for secondary use or spectrum usage modeling is not an easy task in such scenario. The spectrum measurements for a longer period are useful to characterize and model the spectrum usage. The spectrum usage modeling is the important area of research to realize the cognitive radio. It helps in spectrum sensing, the key function of the cognitive radio. The existing research literature has shown the Beta distribution in the modeling the spectrum usage of the licensed user in a licensed band. This article has validated the Beta distribution channel usage model using the real-time measurements that was conducted in the frequency band 700 to 2746.6 MHz at Pune, India. In this paper, Kolmogorov-Smirnov (K-S) Test is used for Beta distribution validation.

General Terms

Spectrum, Measurement

Keywords

Cognitive Radio, Spectrum usage, Beta distribution, Kolmogorov-Smirnov Test

1. INTRODUCTION

Day by day the wireless device users are increasing rapidly and so as their bandwidth requirements. The spectrum regulatory body has divided the radio spectrum among the various wireless services. The spectrum measurement survey has shown the spectrum usage across the radio spectrum band is not uniform [1][2]. It has been observed that some of the bands showed heavy utilization, some moderately and some sparsely utilized. The cognitive radio (CR) [3] [4] is a promising solution for the current discrepancy of spectrum utilization in the radio spectrum. The CR users look for the unused primary user (PU) frequencies for communication purpose and thus improve spectrum utilization and fulfill the user requirement of the spectrum.

The modeling of spectrum usage for the study of wireless communication systems is the crucial step towards its deployment. The primary users' arrival and departure in the licensed band are random in nature, and this poses a problem in modeling the spectrum usage. In CR paradigm, the secondary user opportunistically accesses the unused licensed spectrum, so the spectrum usage pattern of PU system plays significant role in behavior and performance of CR system. Thus, the accurate and realistic spectrum usage modeling of PU system is very much useful in CR technology. The existing spectrum usage models classified as time, frequency and space dimension models in [5]. In time dimension modeling, the two-state Markov chain model is the most widely used in cognitive radio research. The spectrum occupancy model presented in [6] produces both the temporal and frequency behavior of different wireless transmissions and verified by measurements. The existing research literature has shown the Beta distribution in the modeling the spectrum usage of the PU in a licensed band. The author assumed Beta distribution for channel occupancy modeling and corroborated with real-time measurements in [7]. This assumption validated in [8] using real-time measurement performed in Aachen, Germany on the 1500 MHz spectrum centered at 770 MHz. The Beta distribution is a good fit to the channel occupancy due to the high, moderate and low probability of occupancy across the radio spectrum [8]. The continuous broadcast channels e.g. the downlink channels in GSM or CDMA services have a high probability of occupancy. The channels allocated to PMRTS, FM, and ISM has a moderate probability of occupancy. The channels allocated to space navigation and radio astronomy has a low probability of occupancy. M.wellens and P.Mähönen have shown modified beta distribution is a good model for the duty cycle distribution [9]. In this paper, the beta distribution assumption of channel occupancy is validated using our real time measurement performed at Pune, India in the frequency band of 700 to 2746.6 MHz.

The remainder of this paper is structured as follows. Section 2 provides details of the measurement setup. Section 3 presents brief introduction regarding beta distribution. Section 4 provides results of K-S test used for validation of beta distribution. Lastly, section 5 concludes the paper.

2. SPECTRUM MEASUREMENT



Fig 1. Measurement Setup

The measurement setup employed in this campaign shown in figure 1 consisted of spectrum analyzer, a laptop and an antenna. A Rohde and Schwarz (FSH3) spectrum analyzer interfaced with a laptop via an optical cable. The spectrum analyzer was programmed to save real-time data in the laptop's storage device. AOR DA 5000 antenna was used in this campaign. The specified frequency range of operation of an antenna is from 700 MHz to 3GHz. A low loss coaxial cable is used to connect an antenna with a spectrum analyzer. The details of measurement setup, methodology and spectrum analyzer configuration specified in [10]. The setup has been installed on the terrace of a commercial building on law college road, a busy location of FSH using the commands

for spectrum analyzer was used for real-time data collection. The data has been collected continuously for seven days. The files collected were in .csv format. Further, the data has been processed and viewed for analysis and comparison using MATLAB software.

3. BETA DISTRIBUTION

A continuous distribution whose probability density function takes on many different shapes, and is described by two positive shape parameters, is Beta distribution. The PDF of a Beta distribution is given by

$$f_x^B(x,\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1} (1-x)^{(\beta-1)}, \ x \in (0,1)$$
(1)

Where $\alpha > 0$ and $\beta > 0$ are shape parameters and $B(\alpha, \beta)$ is the Beta function. Beta function defined as

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$$
 (2)

The beta function $B(\alpha, \beta)$ is normalization constant. The choice of Beta distribution for generating a set of channel free probabilities randomly because any continuous probability distribution on (0, 1) can be approximated by a Beta distribution [11].

4. VALIDATION OF BETA DISTRIBUTION USING K-S TEST

4.1 Kolmogorov-Smirnov (K-S) test

The Kolmogorov-Smirnov (K-S) test [12] is used for the validation of assumption of the beta distribution. This test is Goodness of Fit test (GoF) based on statistical theory. This approach assesses the underlying distribution of the data set. The K-S test is based on the empirical cumulative distribution function (ECDF). The Kolmogorov-Smirnov statistic (D) represents the largest vertical difference between $F_0(x)$ and $F_n(x)$. Where $F_0(x)$ is assumed cumulative distribution and $F_n(x)$ is empirical cumulative distribution function.

The data set obtained by a long-term measurement is used for validation. The data set was divided into the three or four intervals of interest. The random sample of channels has been taken with occupancy probabilities between 0, and 1 for each time interval under consideration. Then different mean (μ_{usage}), and variance (σ_{usage}) of the channel usage over the different data sets evaluated. To implement the K-S test, well-defined series of steps are followed. First, the distribution is assumed as a beta distribution. Then estimated the α and β i.e. distribution parameters from the data set. The α and β parameters can be computed from the obtained data using the following expression [13]:

$$\hat{\alpha} = \mu_{usage} \left(\mu_{usage} \frac{(1 - \mu_{usage})}{\sigma_{usage}} - 1 \right)$$
(3)
$$\hat{\beta} = \left(1 - \mu_{usage} \right) \left(\mu_{usage} \frac{(1 - \mu_{usage})}{\sigma_{usage}} - 1 \right)$$
(4)

Above mentioned procedure results in a distribution hypothesis, also called null hypothesis H_0 , if many components of it supported by the data. The negation of the assumed distribution is the alternative hypothesis, also called H_1 . The stepwise KS GoF test procedure was used in this work given below.

Step 1: Establish the (H₀) assumed distribution: Beta

Step 2: Estimate the Beta parameters: α , β

Step 3: Sort the data in ascending order

Step 4: Obtain the theoretical distribution: $F_0(x)$

Step 5: Obtain the empirical distribution: $F_n(x)$

Step 6: Obtain D+= F_n - F_0 and D-= F_0 - F_{n-1} for every data point of x

Step 7: Obtain the KS statistics D = max (D+,D-)

Step 8: Obtain the KS table critical value (CV)

Step 9: Since KS < CV, assume the data from a Beta

In this KS test, when the assumed distribution is correct, the assumed cumulative distribution function closely, follows the empirical cumulative distribution function there. The data is given as an ordered sample, and the assumed (H₀) theoretical distribution has a CDF, $F_0(x)$. Then, the corresponding GoF test statistical values are obtained. Lastly, the theoretical and empirical results are compared. Assumed distribution gets confirmed if they match. The distribution assumption is rejected if they do not agree.

Table 1. Estimated Parameters from the K-S test.

Time Interval	μ_{usage}	σ _{usage}	α	β	h	р	k
9-10 AM	0.43	0.14	0.32	0.41	0	0.20	0.10
2-3 PM	0.37	0.12	0.30	0.50	0	0.13	0.11
7-8 PM	0.45	0.14	0.30	0.36	0	0.25	0.10

4.2 Validation of Beta Distribution

The beta distribution assumption is validated for primary user channel occupancy using real-time measurements in a band from 700 to 2746.6 MHz in Pune. The data of the seven days period have been divided into three intervals of interest, as given below.

Morning	9.00-10.00 am					
Afternoon	2.00-3.00 pm					
Evening	7.00-8.00 pm					
T (1 · · · · 1	200 1 1					

In this work, 200 random channels from the measured band were considered. A random sample of channels with occupancy probabilities between 0 and 1 is taken over the



Fig. 2. Beta density and histogram of occupancy probabilities for 9-10 AM



Fig 3. Beta density and histogram of occupancy probabilities for 2-3 PM



Fig 4. Beta density and histogram of occupancy probabilities for 7-8 PM

each time interval. The mean (μ_{usage}), and variance (σ_{usage}), alpha and beta parameters are estimated from the data set over the specified duration of an hour and given in the table 1.The validation of beta distribution has been performed using the K-S test in MATLAB. The null hypothesis is tested using the K-S test. The K-S test compares the distributions of the values in the two data vectors. The null hypothesis is that the two data sets are from the same continuous distribution. The alternative hypothesis is that they are from different continuous distributions. The result h is one if the test rejects the null hypothesis at the five percent significance level; 0 otherwise. Table 1 list out the parameters from the K-S test. It is evident from the h value = 0 the null hypothesis of Beta distribution assumption for spectrum occupancy is accepted based on the K-S test. The observed frequencies and the beta distribution with the estimated parameters are shown in the figure 2, 3 and 4 for different time intervals. It is clear from the figures that the beta distribution is a good fit for the frequency band 700 to 2746.6 MHz under the study.

5. CONCLUSION

The spectrum usage modeling of the primary system is the key aspect of analysis, design, and simulation of the cognitive radio system. Real-time data from longer period measurements is useful to characterize and modeling the use of spectrum. The spectrum measurement campaign performed for seven days at Pune; India is presented. In this work, the assumption of Beta distribution for spectrum usage is validated using real-time measurements conducted in frequency band 700 to 2746.6 MHz. Kolmogorov-Smirnov test is applied for the validation.

6. REFERENCES

- [1] M. Islam, G. L. Tan, F. Chin, B. E. Toh, Y.-C. Liang, C. Wang, Y. Y. Lai, X. Qing, S. W. Oh, C. L. Koh, and W. Toh, "Spectrum Survey in Singapore: Occupancy Measurements and Analyses", in Proc. of International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), Singapore, May 2008, pp. 1-7.
- [2] Kishor Patil, Knud Skouby, Ashok Chandra, Ramjee Prasad, "Spectrum Occupancy Statistics in the Context of Cognitive Radio," in Proc. of 14th International Symposium on Wireless Personal Multimedia Communications (WPMC'11),Brest, France, November 2011.
- [3] J. Mitola, III and G. Q. Maguire, Jr., "Cognitive radio: making software radios more personal," IEEE Personal Communications, Vol. 6, Issue 4, pp. 13-18, Aug. 1999.
- [4] S.Haykin, "Cognitive Radio: Brain-empowered wireless communications," IEEE Journal on Selected Area in Communications, vol. 23, no. 2, pp. 201–220, February 2005.
- [5] Miguel Lopez-Benitez and Fernando Casadevall "An Overview of Spectrum Occupancy Models for Cognitive Radio Networks" Proceedings of the IFIP international workshop on performance evaluation of cognitive radio networks (PE-CRN 2011), 13 May 2011, pp 1–10.
- [6] C. Ghosh, S. Pagadarai, D. Agrawal, and A. Wyglinski, "A framework for statistical wireless spectrum

occupancy modeling," Wireless Communications, IEEE Transactions on, vol. 9, no. 1, pp. 38–44, 2010.

- [7] P. F. Marshall, "Closed-form analysis of spectrum characteristics for cognitive radio performance analysis," IEEE DySpan, Chicago, IL, 2008, pp. 1-12.
- [8] C. Ghosh, S. Roy, M. B. Rao, and D. P. Agrawal, "Spectrum occupancy validation and modeling using real-time measurements," in Proceedings of the 2010 ACM workshop on CR networks, ser. CoRoNet '10, 2010, pp. 25-30.
- [9] M.Wellens and P.Mähönen, "Lessons Learned from an Extensive Spectrum Occupancy Measurement Campaign and a Stochastic Duty Cycle Model," Springer Mobile Networks and Applications, August 2009.
- [10] Kishor Patil, Ramjee Prasad, Knud Skouby, "Stochastic Duty Cycle Model Based on Measurement for Cognitive Radio", in Proc. of 15th International Symposium on Wireless Personal Multimedia Communications (WPMC'12), pp.128-132, Taipei, Taiwan, September 2012.
- [11] P. M. Lee. Bayesian Statistics: An Introduction, Oxford University Press, Third Edition, 2004.
- [12] M. Chakravarti, R. G. Laha and J. Roy, Handbook of Methods of Applied Statistics, Volume I, John Wiley and Sons, pp. 392-394, 1967.
- [13] A. Stuart, K. Ord, and S. Arnold, Kendall's Advanced Theory of Statistics, vol. 2A, Wiley, 6th edition, 1999.