SURVIVAL ANALYSIS OF REPORTED CASES OF DIABETES DISEASE IN NIGERIA

by

Omaku, Peter Enesi,Nafisat Tanko&Yunusa Ibrahim Adamu

Department of Mathematics & Statistics, Federal Polytechnic Nasarawa – Nigeria

Email: omakupete@gmail.com, ibrosta111@gmail.com

Abstract

This work aim to investigate the distribution of survival time for the reported cases of diabetic patients at the Air force base hospital Abuja, evaluating various covariates. The Kaplan Meire-estimator suggested that there is no significance difference in the distribution of survival time by sex as married patients were observed to survive longer than single patients. The estimate of survival distribution upon test for patients in urban and rural areas are seen to be the same. On the Cox proportional model, we see a model that was significant upon test as P-value = (0.000) is seen to be less than the 0.05 threshold. We again see that the relative risk of patients is dependent on age as the distribution of survival time for patients with diabetes is seen to be significantly different for patients of the four age categories considered in the study. Every patient is expected to get the hazard at an approximately the same time with no multiplication effect with respect to sex. We conclude that the prevalence of the disease is independent on some of the covariates consider which arouse the need for more frequent medical examinations.

Keywords:*survival function, hazard function, events, Kaplan-Meire Estimator, proportional hazard, cox regression, covariates, diabetes.*

1. Introduction

Survival Analysis typically focuses on time to event data. In the most general sense, it consists of techniques for positive- valued random variables, such as time to death, time to onset (or relapse) of a disease, Length of stay in a hospital. The term survival time refers to the length of time t, that corresponds to the time period from a well defined start-time t_o until the occurrence of some particular event or end-point t_ci.et = t_c - t_o. (Arbia et al, 2016). In biomedical research, event could be death, remission from a disease, occurrence of an epileptic seizure etc.

An aspect of analysis of survival time data that has gained popularity, especially in medical research is assessing the relationship between survival time and some biological, socio-economic and demographic characteristics that could possibly affect the survival status of patients. Due to censoring, standard linear regression methods are not feasible in modeling such relationship(Singh et al, 2011).Censoring occurs when the actual time a subject experiences the event of interest is not known. Two notable formulations often used are the Kaplan Meier estimator and the Cox regression model.

The Kaplan Meierestimatoris based on individual survival times and assumes that censoring is independent of survival time (that is, the reason an observation is censored is unrelated to the cause of failure).

Kaplan –Meier Estimator (K-M) is a statistical technique used to analysis survival data. It is applied in analyzing the distribution of the patient's survival times following their recruitment into the study. The analysis expresses this in terms of proportion of patients still alive up to a given time following the recruitment or entry into the study. The K-M estimator is also called non parametric maximum likelihood estimator. It is used for estimating survival probabilities. The method computes the probability of dying at a certain point in time conditional on the survival, up to that point when the patient is censored. Thus, it maximizes utilization of available information on time to event of the study sample.

$$St = \frac{\text{Number of individuals surviving longer than time t}}{\text{total number of individuals studied}} \qquad (1.1)$$

t is a time period known as the survival time, time to failure or time to even$t$ (such as death), (Kalbfleisch et al, 2011), (Brostrom et al, 2012).

Cox Proportional Hazard model

The popular regression model formulation that is often used in survival analysis is the Cox (1972) proportional hazards model. The model utilizes the hazard function *h*(*t*), also known as the hazard rate or force of mortality and it is defined as the probability of experiencing event of failure in the infinitesimally small interval (t, $t+\Delta t$), given that such an event has not been experienced prior to *t*. It is expressed as

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\langle t \le T \le t + \Delta t \mid T > t \rangle}{\Delta t} \qquad (1.2)$$

The Cox's proportional hazard model, is a bench mark method in survival and event history data analysis. Cox regression (or proportional hazards regression) is method for investigating the effect of several variables upon the time a specified event takes to happen. In the context of an outcome such as death this is known as Cox regression for survival analysis. The method does not assume any particular "survival model" but it is not truly nonparametric because it does assume that the effects of the predictor variables upon survival are constant over time and are additive in one scale.(Kalbfleisch et al. 2011).(Keinbaum, et al, 2012)

However, its functional form for the dependency of the survival time on the covariates is fully parametric. In other words, the regressors are linearly related to the log hazard. Most often, in modelling survival time data, either the true hazard is not known or it is complex in which case assumption of parametric model may not be true for such data. For example, when a Weibull model is used for analysis of data from a population that is not from a Weibull survival distribution, the Cox model analysis is more efficient than the parametric model analysis (Richard, 2012).Cox model could also be an alternative model that is as efficient as parametric models such as Weibull model with proportional hazard even when all the parametric assumptions are satisfied.

According to literature, diabetes is a chronic disease for which there is no known cure except in very specific situations. Management of this disease, however, concentrates on keeping blood sugar level close to normal, without causing low blood sugar. This can usually be accomplished with a healthy diet, exercise, weight loss, and use of appropriate medications (insulin in the case of type 1 diabetes, oral medications as well possibly insulin, in type 2 diabetes)

In this paper we aim to employ survival analysis technique to assess the reported case of diabetes in Nigeria with some covariates:we however, examine thedistribution of survival time for diabetes patients via their covariates using Kaplan Meire estimator and also fit a cox proportional hazard model.

2. Kaplan Meire Estimator and the Cox Proportional Hazards Model Formulation

2.1 Model Specification of Kaplan Meire estimator

Suppose t_j, j = 1, 2, ..., n is the total set of failure times recorded (with t^+ the maximum failure time), d_j is the number of failures at time t_j , and r_j is the number of individuals at risk at time t_j.

(1) . for each time period the number of individuals present at the start of the period is adjusted according to the number of individuals censored and the number of individuals who experienced the event of interest in the previous time period, and

(2) for ties between failures and censored observations, the failures are assumed to occur first.

$$\hat{S}(t) = \prod_{j:t_j \le t} \left(\frac{r_j - d_j}{r_j}\right), for\ 0 \le t \le t^+ \qquad (2.1)$$

2.1.1 Model Specification of Cox proportional model

Suppose that the data collected on *n* subjects are denoted by (t_i, δ_i, Z_i), where $t_i$is time to failure of the *ith* subject, δ_i is the censoring indicator such that for the *ith* subject, δ_i=1 if a subject is observed to failure and $\delta_i = 0$ if the time is right censored (i.e we observe some value *c* with the knowledge that $t_i > c$)and Z_i is a p- dimensional vector of covariates. Cox (1972) model assumes that the hazard function for the *ith* subject with covariate value Z_i has the form

$$\lambda(t_i, Z_i) = \lambda_0(t)\exp(\beta^1 Z_i) \qquad (2.2)$$

where $\lambda_0(t)$ is an arbitrary baseline hazard function and β is a p- vector of unknown regression coefficients. Model (2.2) is semi-parametric because the dependence function, $\exp(\beta^1 Z_i)$ is modeled explicitly but no specific probability distribution is assumed for the survival times. Thus β is only estimable through the partial likelihood estimation procedure(Richard, 2012).

2.2 Estimation of Parameters

Suppose that of the *n* subjects in the study, *r* of them are observed to fail while the remaining *n-r* are right-censored. Let $t_{(1)} < \ldots < t_{(r)}$ be ordered failure times and $Z_{(i)}$ be the vector of covariates associated with the individual whose survival time is $t_{(i)}$. Define $R(t_{(i)})$, the risk set at $t_{(i)}$ as the set of all individuals who are still under study at the time just prior to $t_{(i)}$, then the probability that the individual with covariate $Z_{(i)}$ dies at $t_{(i)}$ given that one person from $R(t_{(i)})$ dies at $t_{(i)}$ is

$$\frac{\lambda(t_{(i)}, Z_{(i)})}{\sum_{j \in R(t_{(i)})} \lambda(t_{(i)}, Z_{(i)})} \qquad (2.3)$$

which from (2.2), is

$$\frac{\exp(\beta^1 Z_{(i)})}{\sum_{j \in R(t_{(i)})} \exp(\beta^1 Z_{(i)})} \qquad (2.4)$$

Cox (1972), on the assumption of no tied events, gave the partial likelihood function as

$$L(\beta) = \prod_{i=1}^{r} \left[\frac{\exp(\beta^1 Z_{(i)})}{\sum_{j \in R(t_{(i)})} \exp(\beta^1 Z_j)} \right] \qquad (2.5)$$

and the log partial likelihood is

$$\log L(\beta) = \sum_{i=1}^{r} \left[\beta^{\prime} Z_{(i)} - \log \sum_{j \in R(t_{(i)})} \exp(\beta^{\prime} Z_j) \right] \qquad (2.6)$$

Often, ties occur in continuous survival data that are collected in days, weeks and months. When there are only a few ties, (Chen et al,2014)provided an approximation to (2.4) as

$$L(\beta) = \prod_{i=1}^{r} \left[\frac{\prod_{j \in D(t_{(i)})} \exp(\beta^{\prime} Z_j)}{\left(\sum_{j \in R(t_{(i)})} \exp(\beta^{\prime} Z_j) \right)^{d_i}} \right] \qquad (2.7)$$

where $R(t_{(i)})$ and $t_{(i)}$ are as earlier defined, $D(t_{(i)})$ is the set of individual failing at $t_{(i)}$ and d_i is the number of failures occurring at $t_{(i)}$.

The log likelihood of (2.6) is

$$\log L(\beta) = \sum_{i=1}^{r} \left[\sum_{j \in D(t_{(i)})} \beta^{\prime} Z_j - d_i \log \sum_{j \in R(t_{(i)})} \exp(\beta^{\prime} Z_j) \right] \qquad (2.8)$$

Efron (1977) also derived likelihood that is an improvement over (2.6) given as

$$L(\beta) = \prod_{i=1}^{r} \left[\frac{\prod_{j \in D(t_{(i)})} \exp(\beta^{\prime} Z_j)}{\prod_{k=1}^{d_i} \left(\sum_{j \in R(t_{(i)})} \exp(\beta^{\prime} Z_j) - \frac{k-1}{d_i} \sum_{j \in D(t_{(i)})} \exp(\beta^{\prime} Z_j) \right)} \right] \qquad (2.9)$$

with log likelihood

$$\log L(\beta) = \sum_{i=1}^{r} \left[\sum_{j \in D(t_{(i)})} \beta^{\prime} Z_j - \sum_{k=1}^{d_i} \log \left(\sum_{j \in R(t_{(i)})} \exp(\beta^{\prime} Z_j) - \frac{k-1}{d_i} \sum_{j \in D(t_{(i)})} \exp(\beta^{\prime} Z_j) \right) \right] \qquad (2.10)$$

The maximum partial likelihood estimators $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ can be obtained for (2.6) and (2.10) from the solution of the estimating equation involving the score statistics

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta_k} = 0, \quad k = 1, \ldots, p \qquad (2.11)$$

and the information matrix can be obtained from

$$\mathrm{I}(\beta) = -\frac{\partial^2 \log L(\beta)}{\partial \beta_k \partial \beta_h} \qquad (2.12)$$

Using (2.13) and (2.14), $\hat{\beta}$ can be obtained by solving the iterative equation

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{m} + \mathrm{I}^{-1}(\hat{\beta}^{m}) U(\hat{\beta}^{m}) \qquad (2.13)$$

However, through the exponential link function, the covariates act multiplicatively on the hazard rate. In the case of time-constant covariates, the influence of the covariates implies that the hazard rates for any two individuals are proportional, which explains why the Cox model is called a proportional hazards model. Suppose that Z_i and $Z_j$denote the covariate vectors of two individuals i and j, then the ratio of the hazard rates of these individuals is given by

$$\frac{\lambda(t_i, Z_i)}{\lambda(t_j, Z_j)} = \exp(\beta^{\prime} Z_i - \beta^{\prime} Z_j) \qquad (2.14)$$

$$\lambda(t|Z_i) = c.\lambda(t|Z_j) \qquad (2.15)$$

Where c = $\exp(\beta^{\prime} Z_i - \beta^{\prime} Z_j)$

3. Analysis of Diabetes Data

In this section, we analyze the survival times of 453 diabetic patients who were admitted at the Nigerian air force base hospital Abuja. Time from diagnosis of the disease to death defines the failure time while those whose records read "alive "were right-censored because such patients had not died as at the time of the study.

About 45% of the patients were censored, which on the average, is an indication that hospital admissions of patients would possibly result in eventual death due to diabetes.

Several covariates sex and age, marital status, blood sugar level, diabetes type, area at the time of diagnosis were considered. Sex was coded 1 for male and 0 for female patients. Age was coded into four groups: < 23 years =1, 23-39 years = 2, 40-55 years = 3 and > 55 years = 4. Marital Status was coded 1 for married and 0 for Single. Area was coded 1 for patients from urban and 0 for patients from rural. The median times of hospital admission before death are computed for male and female patients, for the four age groups, for married and single patients. The log-rank statistic, which is a Chi-square type statistic, is also computed to test the equality of survivals between male and female patients and among the age groups. Analyses are also carried out using (2.1) and (2.2) within the framework of (2.7) and (2.12). All analyses have been done using SPSS 16 and the results are presented in section 4.

4. Results

Kaplan-Meier estimator for male and female patients with diabetes

Table 1.Means and Medians for Survival Time for Male and Female Patient with Diabetes

Gender, Male = 1, Female = 0	Mean[a]	Median	Log Rank Mantel-COX	
	Estimate	Estimate	0.968	P-value=0.325
0	4.070	4.300		
1	3.893	3.500		
Overall	3.950	4.000		

Table 1 shows that the female patients survive a little longer than the male patients with mean survival time of 4.070 years compared to 3.893 years their male counterpart. Same is also observe in the median survival time as female survive longer at a medianyears of 4.300 against the male with median time of 3.500 years. The log-rank test statistic $\chi^2 = 0.968\,(p = 0.325)$ is an indication that there is no significant difference between male and female patients.

Kaplan-Meier estimator for married and single patients with diabetes

Table 2.Means and Medians for Survival Time for Male and Female Patient with Diabetes

M.Status, Married = 1, single = 0	Mean[a]	Median	Log Rank Mantel-COX	
	Estimate	Estimate	20.703	P-value=0.000
0	3.085	3.200		
1	4.140	4.210		
Overall	3.950	4.000		

Table 2 show that married patients survive longer than the single patients (reference category) with a mean of survival time of 4.210 years compare to 3.200 is for reference category. Same is also observed in median survival time (as married patients survive longer at 4.120 years as against the single patients with median time of 3.200. The log-rank test statistic $\chi^2 = 20.703\,(p = 0.000)$ is an indication of significant difference between male and female patients.

Kaplan Meire Estimator for patients in urban and rural areas.

Table 3.Means and Medians for survival time for urban and rural area patients with diabetes

Area, urban = 1, rural = 0	Mean[a]	Median	Log Rank Mantel-COX	
	Estimate	Estimate	0.472	P-value=0.492
0	4.014	4.100		
1	3.881	3.500		
Overall	3.950	4.000		

Table 3 show that patients in rural areas survive longer that the urban area patients (reference category) with a mean survival time of 4.014 years as compare to 3.881 years for those in rural area. Same is also observed in medial survival time of (as rural patients survive longer at median time of 4.100 years as against the urban with median of 3.500 years. The log-rank test statistic $\chi^2 = 0.472(p = 0.492)$ is an indication that there is no significant difference between male and female patients.

Cox Regression

Table 4 Summary of Cox Proportional Hazards Model

Omnibus Tests of Model Coefficients[a,b]									
-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
2481.763	98.359	7	.000	76.428	7	.000	76.428	7	.000

From table (4) the partial likelihood produce significant model for the diabetes data with LR tests 98.359 (p=0.000).

Table 5: Table of coefficients and hazard rates

Variables in the Equation						
	B	SE	Wald	df	Sig.	Exp(B)
Sex	.350	.139	6.395	1	.011	1.419
MStatus	-.145	.180	.652	1	.419	.865
Area	-.050	.130	.149	1	.700	.951
BSL	.158	.096	2.690	1	.101	1.171
Grp4	2.264	.333	46.277	1	.000	9.623
Grp3	1.673	.307	29.777	1	.000	5.328
Grp2	.874	.290	9.044	1	.003	2.396
Grp1			.	0[a]	.	

From table 5: The estimated relative risks for male versus female patients is 1.419, meaning that the risk of dying from diabetes by male patients is 1.419 times that of female. For Married versus single patients is 0.869, meaning that the risk of dying from diabetes by married patients is 0.869 times that of single patients, for instance for every 100 married patients we have 87 more deaths of single patients as a result of diabetes. For urban versus rural patients is 0.951, meaning that the risk of dying from diabetes by urban patients is 0.951 times that of rural patients

For age groups, the risk of dying from diabetes for patients in age groups 23-39 years, 40-55 years and >55 years relative to the baseline age group (<23 years) are 2.396, 5.328 and 9.623 respectively. Which implies that patients that are older are at risk of death from diabetes than younger patients.

5. Conclusion

Comparing the time of Hospital admission before death, based on the diabetes data, male diabetic patients are seen to have an approximately equal survival time with their female counterparts, married patients survive diabetes longer than single patients, this may be pin down to diet control that married patient may have against diabetic patients that are single. The estimate of survival distribution upon test for patients in urban and rural areas are seen to be the same, we however observe that those in rural area survive a little longer than those in urban areas.

Also from the relative risk point of view, relative risk (hazard ratio) of male versus female (baseline) having value greater than one (1) implies higher risk of dying from diabetes by male than female patients which is an indication of superior survival for female than male patients. For age group, the risk of dying from diabetes increases progressively with age.

References

Arbia G, Espa G, Giuliani D, Micciolo (2016). "A Spatial Analysis of Health and Pharmaceutical Firm Survival." Journal of Applied Statistics, p. in press.

Brostrom, Goran (2012), Event History Analysis with R (First ed.), Chapman and HALL/CRC, ISBN 978-1584884088

Chen Y, Hanson T, Zhang J (2014). "Accelerated Hazards Model Based on Parametric Families Generalized With Bernstein Polynomials." Biometrics, 70(1), 192–201.

Cox, D.R., (1972): Regression models and life tables (with discussions). Journal of the Royal Statistical Society, B, 34: 187-220

Kalbfleisch, J. D. and Prentice, R. L. (2011).The statistical analysis of failure time data, volume 360. John Wiley & Sons.

Keinbaum, David. G; Klein,Mitchel (2012), Survival analysis: A self-learning text (third ed.),Springer, ISBN 978-1441966452

Richard, S.J. (2012). A handbook of parametric survival models for actuarial use. Scandinavian Actuarial Journal. 2012(4) 233-257 doi:10.4103/2229-3485.86872.

Singh, R.; Mukhopadhyay, K.(2011). "Survival analysis in clinical trials: Basic and must know areas". PerspectClin Res. 2(4): 145-148. Doi:10.4103/2229-3485.86872

AN INTRODUCTION OF THE CONCEPT OF N-LEVEL SOFT SET

by

A. I. Isah

Department of Mathematical Sciences Kaduna State University

ahmed.isah@kasu.edu.ng

Abstract

The theory of soft sets and soft multisets are considered as useful tools for modeling uncertainty. In this paper, the concept of n-level soft set is introduced together with some of its properties. Both the first and the second decomposition theorems were established and proved.

Key words***: Soft set, Multiset, Soft multiset, n-level Soft set***

1. Introduction

The theory of Soft set which was initiated with the aim of modeling uncertainty in real life situation has many applications in areas of decision making, medical diagnosis, data analysis, forecasting, game theory etc. as presented in [1, 2, 3, 4].

By violating a basic underlying set condition, the concept of multiset (mset, for short) which is an unordered collection of objects where multiples of objects are admitted was initiated with the aim of addressing repetition which is significant in real life situations. For a comprehensive account of the idea of multiset and its applications refer to [5, 6, 7, 8, 9].

Soft multiset which is a mapping from a set of parameters to the power set of a universal multiset was studied in different ways as can seen in [10, 11, 12, 13]. However, as multisets are generalization of sets [14], the idea of [13] serves as a generalization of soft sets.

The concept of n-level sets was introduced in [15] and studied by [16] together with some of their properties. In this paper the concept of n-level soft set is introduced and some related results were obtained.

2.1 Soft set

Definition 2.1.1 [17]

Let U be an initial universe set and E a set of parameters or attributes with respect to U. Let $P(U)$ denote the power set of U and $A \subseteq E$. A pair (F, A) is called a *soft set* over U, where F is a mapping given by $F: A \longrightarrow P(U)$.

In other words, a soft set (F, A) over U is a parameterized family of subsets of U. For $e \in A$, $F(A)$ may be considered as the set of e-elements or e-approximate elements of the soft set (F, A). Thus (F, A) is defined as

$$(F, A) = \{F(e) \in P(U): e \in E, F(e) = \emptyset \text{ if } e \notin A\}.$$

Definition 2.1.2 [4]

Let (F, A) and (G, B) be two soft sets over a common universe U, we say that

(a) (F,A) is a **soft subset** of (G,B), denoted $(F, A) \tilde{\subseteq} (G, B)$, if

(i) $A \subseteq B$, and

(ii) $\forall\ e \in A, F(e) \tilde{\subseteq} G(e)$.

(b) (F, A) is **soft equal** to (G, B), denoted $(F, A) = (G, B)$, if $(F, A) \tilde{\subseteq} (G, B)$ and $(G, B) \tilde{\subseteq} (F, A)$.

Definitions 2.1.3 [4, 18]

Let (F,A) and (G,B) be two soft sets over a common universe U.

(i) The **union** of (F,A) and (G,B), denoted $(F,A)\tilde{\cup}(G,B)$, is a soft set (H,C) where $C = A \cup B$ and $\forall\ e \in C$,

$$H(e) = \begin{cases} F(e), & e \in A - B \\ G(e), & e \in B - A \\ F(e) \cup G(e), & e \in A \cap B. \end{cases}$$

(ii) The **extended intersection** of (F,A) and (G,B), denoted $(F,A)\tilde{\cap}(G,B)$, is a soft set (H,C) where $C = A \cup B$ and $\forall\ e \in C$,

$$H(e) = \begin{cases} F(e), \text{ if } e \in A - B \\ G(e), \text{ if } e \in B - A \\ F(e) \cap G(e), \text{ if } e \in A \cap B. \end{cases}$$

(iii) The **restricted intersection** of (F,A) and (G,B), denoted $(F,A)\cap_R (G,B)$, is a soft set (H,C) where $C = A \cap B$ and $\forall\ e \in C$, $H(e) = F(e) \cap G(e)$. If $A \cap B = \phi$ then $(F,A)\cap_R (G,B)= \tilde{\Phi}_\phi$.

(iv) The **restricted union** of (F,A) and (G,B), denoted $(F,A)\cup_R (G,B)$, is a soft set (H,C) where $C = A \cap B$ and $\forall\ e \in C$, $H(e) = F(e) \cup G(e)$. If $A \cap B = \phi$ then $(F,A)\cup_R (G,B) = \tilde{\Phi}_\phi$.

2.2 Multisets

Definition 2.2.1 [19]

An mset M drawn from the set X is represented by a function *Count M* or C_M defined as $C_M: X \longrightarrow \mathbb{N}$. Let M be a multiset from X with x appearing n times in M. It is denoted by $x \in^n M$. $M = \{k_1/x_1, k_2/x_2, \ldots, k_n/x_n\}$ where M is a multiset with x_1 appearing k_1 times, x_2 appearing k_2 times and so on.

Let M and N be two msets drawn from a set X. Then

$M \subseteq N$ iff $C_M(x) \le C_N(x)$ for all $x \in X$.

$M = N$ if $C_M(x) = C_N(x)$ for all $x \in X$.

$M \cup N = max\{C_M(x), C_N(x)\}$ for all $x \in X$.

$M \cap N = min\{C_M(x), C_N(x)\}$ for all $x \in X$.

$M - N = max\{C_M(x) - C_N(x), 0\}$ for all $x \in X$.

Definition 2.2.2 [20]

Let M be a multiset drawn from a set X. The support set of M denoted by M^* is a subset of X given by $M^* = \{x \in X: C_M(x) > 0\}$. Note that $M \subseteq N$ iff $M^* \subseteq N^*$.

The power multiset of a given mset M, denoted by $P(M)$ is the multiset of all submultisets of M, and the power set of a multiset M is the support set of $P(M)$, denoted by $P^*(M)$.

Example 2.2.1 Let $M = \{2/x, 2/y\}$, then $M^* = \{x, y\}$ and $P(M) = \{\emptyset, 2/\{1/x\}, 2/\{1/y\}, \{2/x\}, \{2/y\}, 4/\{1/x, 1/y\}, 2/\{2/x, 1/y\}, 2/\{1/x, 2/y\}, \{2/x, 2/y\}$. Moreover, $P^*(M) = \{\emptyset, \{1/x\}, \{1/y\}, \{2/x\}, \{2/y\}, \{1/x, 1/y\}, \{2/x, 1/y\}, \{1/x, 2/y\}, \{2/x, 2/y\}\}$.

Definition 2.2.4 [20]

Let $\{M_i : i \in I\}$ be a nonempty family of msets drawn from a set X. Then

(i) Their Intersection, denoted by $\bigcap_{i\in I} M_i$ is defined as

$$C_{\bigcap_{i\in I} M_i}(x) = \bigwedge_{i\in I} C_{M_i}(x), \forall x \in X,$$

where $\wedge$ is the minimum operation.

(ii) Their Union, denoted by $\bigcup_{i\in I} M_i$ is defined as

$$C_{\bigcup_{i\in I} M_i}(x) = \bigvee_{i\in I} C_{M_i}(x), \forall x \in X,$$

where $\vee$ is the maximum operation.

2.3 Soft Multiset

Definition 2.3.1 [13]

Let U be a universal multiset, E be a set of parameters and $A \subseteq E$. Then a pair (F, A) is called a soft multiset where F is a mapping given by $F : A \longrightarrow P^*(U)$. For all $e \in A$, the mset $F(e)$ is represented by a count function $C_{F(e)}: U^* \longrightarrow \mathbb{N}$.

Example 2.3.1 Let the universal mset $U = \{3/w, 2/x, 4/y, 1/z\}$, the parameter set $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$, $A = \{e_1, e_2, e_3\}$ and the mapping $F : A \longrightarrow P^*(U)$ be defined as $F(e_1) = \{2/w, 1/y, 1/z\}, F(e_2) = \{1/w, 2/x, 3/y\}$ and $F(e_3) = \{1/x, 2/y\}$. That is, (F, A) is a soft multiset such that for all $e \in A$, the multiset $F(e)$ is represented by a count function $C_{F(e)}: U^* \longrightarrow \mathbb{N}$ as

$$C_{F(e_1)}(w) = 2, \quad C_{F(e_1)}(x) = 0, \quad C_{F(e_1)}(y) = 1, \quad C_{F(e_1)}(z) = 1$$
$$C_{F(e_2)}(w) = 1, \quad C_{F(e_2)}(x) = 2, \quad C_{F(e_2)}(y) = 3, \quad C_{F(e_2)}(z) = 0$$
$$C_{F(e_3)}(w) = 0, \quad C_{F(e_3)}(x) = 1, \quad C_{F(e_3)}(y) = 2, \quad C_{F(e_3)}(z) = 0$$

Thus, $(F, A) = \{ (e_1, \{2/w, 1/y, 1/z\}), (e_2, \{1/w, 2/x, 3/y\}), (e_3, \{1/x, 2/y\})\}$.

Definition 2.3.3 [13]

Let (F, A) and (G, B) be two soft multisets over U. Then

(a) (F, A) is a soft submultiset of (G, B) written $(F, A) \sqsubseteq (G, B)$ if

i. $A \subseteq B$

ii. $C_{F(e)}(x) \leq C_{G(e)}(x), \forall x \in U^*, \forall e \in A$.

(b) $(F, A) = (G, B) \Leftrightarrow (F, A) \sqsubseteq (G, B)$ and $(G, B) \sqsubseteq (F, A)$.

Also, if $(F, A) \sqsubset (G, B)$ and $(F, A) \neq (G, B)$ then (F, A) is called a proper soft submset of (G, B) and (F, A) is a whole soft submset of (G, B) if $C_{F(e)}(x) = C_{G(e)}(x), \forall x \in F(e)$.

(c) **Union:**

$(F, A) \sqcup (G, B) = (H, C)$ where $C = A \cup B$ and $C_{H(e)}(x) = max\{C_{F(e)}(x), C_{G(e)}(x)\}, \forall e \in C, \forall x \in U^*$.

(d) **Intersection:**

$(F, A) \sqcap (G, B) = (H, C)$ where $C = A \cap B$ and $C_{H(e)}(x) = min\{C_{F(e)}(x), C_{G(e)}(x)\}, \forall e \in C, \forall x \in U^*$.

(e) **Difference:**

$(F, E) \backslash (G, E) = (H, E)$ where $C_{H(e)}(x) = max\{C_{F(e)}(x) - C_{G(e)}(x), 0\}, \forall x \in U^*$.

(e) **Null:**

A soft multiset (F, A) is called a Null soft multiset denoted by Φ, if $\forall e \in A \; F(e) = \emptyset$.

(f) **Complement:**

The complement of a soft multiset (F,A), denoted by $(F,A)^c$, is defined by $(F,A)^c = (F^c,A)$ where $F^c: A \longrightarrow P^*(U)$ is a mapping given by $F^c(e) = U \backslash F(e), \forall e \in A$ where $C_{F^c(e)}(x) = C_U(x) - C_{F(e)}(x),\ \forall x \in U^*$.

3. n-Level Soft Set

Definition 3.1

Let (F,A) be a Soft multiset over a universal multiset U and a set of parameters E. Then, we define the *n-level soft set* of (F,A), denoted $(F,A)_n$ as

$$(F,A)_n = \{(e,\{x\}) \mid C_{F(e)}(x) \geq n, n \in \mathbb{N}, \forall e \in A, \forall x \in U^*\}.$$

Example 3.1

Let $E = \{e_1,e_2,e_3,e_4,e_5,e_6,e_7\}$, $A = \{e_1,e_2,e_3\}, B = \{e_1,e_2\}, U = \{7/x, 4/y, 3/z\}$,
$(F,A) = \{(e_1,\{2/x,1/y,3/z\}),(e_2,\{4/x,3/y\}),(e_3,\{1/x\})\}$ and
$(G,B) = \{(e_1,\{2/x,2/z\}),(e_2,\{4/x,3/y\})\}$. Then,

$$(F,A)_1 = \{(e_1,\{x,y,z\}),(e_2,\{x,y\}),(e_3,\{x\})\}$$
$$(F,A)_2 = \{(e_1,\{x,z\}),(e_2,\{x,y\})\}$$
$$(F,A)_3 = \{(e_1,\{z\}),(e_2,\{x,y\})\}$$
$$(F,A)_4 = \{(e_2,\{x\})\}$$
$$(F,A)_n = \Phi, n \geq 5$$

and

$$(G,B)_1 = \{(e_1,\{x,z\}),(e_2,\{x,y\})\}$$
$$(G,B)_2 = \{(e_1,\{x,z\}),(e_2,\{x,y\})\}$$
$$(G,B)_3 = \{(e_2,\{x,y\})\}$$
$$(G,B)_4 = \{(e_2,\{x\})\}$$
$$(G,B)_n = \Phi, n \geq 5.$$

Definition 3.2

Let $(F,A)_n$ be the *n-level soft set* of (F,A), then

$$F_n(e) = \{x \in U^* \mid C_{F(e)}(x) \geq n, n \in \mathbb{N}, \forall e \in A\}.$$

Example 3.2

Observe that, from example 3.1, $F_1(e_1) = \{x,y,z\}, F_1(e_2) = \{x,y\}, F_1(e_3) = \{x\}$.

Theorem 3.1

Let (F,A) and (G,B) be Soft multisets over U and E, suppose $m,n \in \mathbb{N}$. Then,
(i) $((F,A) \sqcup (G,B))_n = (F,A)_n \sqcup (G,B)_n$,
(ii) $((F,A) \sqcap (G,B))_n = (F,A)_n \sqcap (G,B)_n$,
(iii) IF $(G,B) \sqsubseteq (F,A)$ then $(G,B)_n \sqsubseteq (F,A)_n$,
(iv) IF $m \leq n$ then $(F,A)_n \sqsubseteq (F,A)_m$,
(v) $(G,B) = (F,A)$ iff $(G,B)_n = (F,A)_n, \forall e \in A, \forall x \in U^*$.

Proof

(i) Let $x \in ((F,A) \sqcup (G,B))_n \Rightarrow x \in (F,A) \sqcup (G,B),\ C_{F(e)}(x) \geq n,\ C_{G(e)}(x) \geq n$
$\Rightarrow x \in (F,A), C_{F(e)}(x) \geq n$ or $x \in (G,B),\ C_{G(e)}(x) \geq n$
$\Rightarrow x \in (F,A)_n$ or $x \in (G,B)_n$
$\Rightarrow x \in ((F,A)_n \sqcup (G,B))_n$
i.e., $((F,A) \sqcup (G,B))_n \sqsubseteq (F,A)_n \sqcup (G,B)_n$ …(1)
Conversely, let $x \in (F,A)_n \sqcup (G,B)_n$
$\Rightarrow x \in (F,A)_n$ or $x \in (G,B)_n$
$\Rightarrow C_{F(e)}(x) \geq n, \forall e \in A$ or $C_{G(e)}(x) \geq n, \forall e \in B$

$\Longrightarrow x \in (F,A), C_{F(e)}(x) \geq n$ or $x \in (G,B), C_{G(e)}(x) \geq n$
$\Longrightarrow x \in (F,A)$ or $x \in (G,B), C_{F(e)}(x) \geq n, C_{G(e)}(x) \geq n$
$\Longrightarrow x \in (F,A) \sqcup (G,B), C_{F(e)}(x) \geq n, C_{G(e)}(x) \geq n$
$\Longrightarrow x \in \big((F,A) \sqcup (G,B)\big)_n$
i.e.,$(F,A)_n \sqcup (G,B)_n \sqsubseteq ((F,A) \sqcup (G,B))_n$ …(2)
From (1) and (2) the result follows.

Similarly for (ii).
(iii) Let $(G,B) \sqsubseteq (F,A)$ and suppose $x \in (G,B)_n$
$\Longrightarrow C_{G(e)}(x) \geq n, \forall e \in B$
Since $C_{G(e)}(x) \leq C_{F(e)}(x), \forall e \in B$ and $B \subseteq A$, we have $C_{F(e)}(x) \geq n, \forall e \in A$
$\therefore x \in (F,A)_n$
i.e., $(G,B)_n \sqsubseteq (F,A)_n$.
(iv) Let $m \leq n$ and suppose $x \in (F,A)_n$
$\Longrightarrow C_{F(e)}(x) \geq n, \forall e \in A$
$\Longrightarrow C_{F(e)}(x) \geq m, \forall e \in A$
$\Longrightarrow x \in (F,A)_m$
i.e., $(F,A)_n \sqsubseteq (F,A)_m$.
(v) Let $(G,B) = (F,A) \Longrightarrow A = B$ and $C_{F(e)}(x) = C_{G(e)}(x), \forall e \in A, \forall x \in U^*$
$\Longrightarrow \forall n \in \mathbb{N}$, if $C_{F(e)}(x) \geq n$ it imply $C_{G(e)}(x) \geq n, \forall e \in A$ and vice versa
Thus, $(G,B)_n = (F,A)_n$.
Conversely,
Let $(G,B)_n = (F,A)_n \Longrightarrow C_{F(e)}(x) \geq n, \forall e \in A$ and $C_{G(e)}(x) \geq n, \forall e \in B$
$\Longrightarrow C_{F(e)}(x) \geq n,$ and $C_{G(e)}(x) \geq n, \forall e \in A$
$\Longrightarrow C_{F(e)}(x) = C_{G(e)}(x), \forall e \in A$
$\Longrightarrow (G,B) = (F,A)$.

Definition 3.3
Let $S(U,E)$ be the class of all Soft multisets over U and E i.e. $S(U,A) = \{F: A \longrightarrow P^*(U), A \subseteq E\}$. Let $Q \subseteq U^*$, then, we define a soft multiset ${}_n(F,A) \in S(U,A)$ as

$$_n(F,A) = \{(e, nQ) | C_{nQ}(x) = n, \forall e \in A, \forall n \in \mathbb{N}.$$

Example 3.3
Let $Q = \{x,y\}$, then

$$_1(F,A) = \{(e,\{x,y\})\},\ _2(F,A) = \{(e,\{2/x,2/y\})\},\ _3(F,A) = \{(e,\{3/x,/3y\})\}, \dots,\ _n(F,A) = \{(e,\{n/x,n/y\})\}, \forall e \in A, \forall n \in \mathbb{N}.$$

Theorem 3.2 (First Decomposition Theorem)
Let $(F,A)_n$ be a n-level soft set of a soft multiset (F,A), over U and E. Then, $C_{(F,A)}(x) = C_{F(e)}(x), \forall e \in A = \sum_{n \in \mathbb{N}} \mathcal{X}(F_n(e))(x), \forall e \in A = \sum_{n \in \mathbb{N}} \mathcal{X}(F,A)_n(x)$ where $\mathcal{X}(F_n(e))$ is the characteristic function of $\big(F_n(e)\big), \forall e \in A$ and $\mathcal{X}(F,A)_n$ is the characteristic function of $(F,A)_n$.

Proof
Let $x \in F_r(e), \forall e \in A, r = r_1, r_2, \dots, r_m, m = cad(A)$ for $x \in U^*$. Observe that $x \notin F_{r+n}(e), n \in \mathbb{N}$. Then $C_{(F,A)}(x) = C_{F(e)}(x) = r, \forall e \in A$. Now

$$\sum_{n \in \mathbb{N}} \mathcal{X}(F,A)_n(x) = \sum_{n \in \mathbb{N}} \mathcal{X}(F_n(e))(x), \forall e \in A = \sum_{n=1}^{r} \mathcal{X}\big(F_n(e)\big)(x) + \sum_{n \in \mathbb{N}} \mathcal{X}(F_{r+n}(e))(x), \forall e \in A$$
$$= [1 + 1 + \cdots r \text{ times}] + [0 + 0 + \cdots] = \text{r}, \forall e \in A.$$

Hence, $C_{(F,A)}(x) = \sum_{n\in\mathbb{N}} \mathcal{X}(F,A)_n(x)$.

Example 3.4

Consider $(F,A) = \{(e_1, \{2/x, 1/y, 3/z\}), (e_2, \{4/x, 3/y\}), (e_3, \{1/x\})\}$.

Now

$$C_{(F,A)}(x) = \{C_{F(e_1)}(x) + C_{F(e_2)}(x) + C_{F(e_3)}(x)\} = 2 + 4 + 1 = 7.$$

But

$$\mathcal{X}(F_1(e_1))(x) = 1, \mathcal{X}(F_2(e_1))(x) = 1, \mathcal{X}(F_n(e_1))(x) = 0, n \geq 3,$$

$\mathcal{X}(F_1(e_2))(x) = 1, \mathcal{X}(F_2(e_2))(x) = 1, \mathcal{X}(F_3(e_2))(x) = 1, \mathcal{X}(F_4(e_2))(x) = 1, \mathcal{X}(F_n(e_2))(x) = 0, n \geq 5, \mathcal{X}(F_1(e_3))(x) = 1, \mathcal{X}(F_n(e_3))(x) = 0, n \geq 2.$

and thus,

$$\sum_{n\in\mathbb{N}} \mathcal{X}(F,A)_n(x) = \sum_{n\in\mathbb{N}} \mathcal{X}(F_n(e))(x), \forall e \in A = \sum_{n=1}^{2} \mathcal{X}(F_n(e_1))(x) + \sum_{n=1}^{4} \mathcal{X}(F_n(e_2))(x) + \mathcal{X}(F_1(e_3))(x)$$
$$= \mathcal{X}(F_1(e_1))(x) + \mathcal{X}(F_2(e_1))(x) + \mathcal{X}(F_1(e_2))(x) + \mathcal{X}(F_2(e_2))(x) + \mathcal{X}(F_3(e_2))(x)$$
$$+ \mathcal{X}(F_4(e_2))(x) + \mathcal{X}(F_1(e_3))(x) = 1 + 1 + 1 + 1 + 1 + 1 + 1 = 7$$

Theorem 3.3 (Second Decomposition Theorem)

Let $(F,A)_n$ be the n-level soft set of a soft multiset (F,A) over U and E. Then $(F,A) = \coprod_{n\in\mathbb{N}} {}_n(F,A)_n$ where $\sqcup$ is the soft multiset union.

Proof

Let $x \in U^*$ and $C_{(F,A)}(x) = t, \forall e \in A$. This imply that $x \in (F,A)_n$, for $n = 1,2,\ldots,t$ and $x \notin (F,A)_n, \forall n \geq t+1, \ \forall e \in A$.

Now,

$$C_{\left(\coprod_{n\in\mathbb{N}} {}_n(F,A)_n\right)}(x) = \coprod_{n\in\mathbb{N}} {}_n(F,A)_n\,(x)$$
$$= {}_1(F,A)_1 \sqcup {}_2(F,A)_2 \sqcup \ldots \sqcup {}_t(F,A)_t \sqcup_{t+1}(F,A)_{t+1} \sqcup \ldots$$
$$= \cup\{1,2,\ldots,t,0,0,\ldots\} = t, \forall e \in A.$$
$$= C_{(F,A)}(x), \forall e \in A, \forall x \in U^*$$
$$= (F,A).$$

Therefore, $(F,A) = \coprod_{n\in\mathbb{N}} {}_n(F,A)_n$.

Example 3.5

Let $(F,A) = \{(e_1, \{2/x, 1/y, 3/z\}), (e_2, \{4/x, 3/y\}), (e_3, \{1/x\})\}$, we have

$$(F,A)_1 = \{(e_1, \{x,y,z\}), (e_2, \{x,y\}), (e_3, \{x\})\}$$
$$(F,A)_2 = \{(e_1, \{x,z\}), (e_2, \{x,y\})\}$$
$$(F,A)_3 = \{(e_1, \{z\}), (e_2, \{x,y\})\}$$
$$(F,A)_4 = \{(e_2, \{x\})\}$$
$$(F,A)_n = \Phi, n \geq 5$$

and thus,

$${}_1(F,A)_1 = \{(e_1, \{1/x, 1/y, 1/z\}), (e_2, \{1/x, 1/y\}), (e_3, \{1/x\})\}$$
$${}_2(F,A)_2 = \{(e_1, \{2/x, 2/z\}), (e_2, \{2/x, 2/y\})\}$$
$${}_3(F,A)_3 = \{(e_1, \{3/z\}), (e_2, \{3/x, 3/y\})\}$$
$${}_4(F,A)_4 = \{(e_2, \{4/x\})\}.$$

Now,

$${}_1(F,A)_1 \sqcup {}_2(F,A)_2 \sqcup {}_3(F,A)_3 \sqcup {}_4(F,A)_4 \sqcup {}_5(F,A)_5 \sqcup {}_6(F,A)_6 \sqcup \ldots$$
$$= \{(e_1, \{1/x, 1/y, 1/z\}), (e_2, \{1/x, 1/y\}), (e_3, \{1/x\})\} \sqcup \{(e_1, \{2/x, 2/z\}), (e_2, \{2/x, 2/y\})\}$$
$$\sqcup \{(e_1, \{3/z\}), (e_2, \{3/x, 3/y\})\} \sqcup \{(e_2, \{4/x\})\} \sqcup \Phi \sqcup \Phi \sqcup \ldots$$
$$= \{(e_1, \{2/x, 1/y, 3/z\}), (e_2, \{4/x, 3/y\}), (e_3, \{1/x\})\} = (F,A).$$

Conclusion

In this paper, the notion of n-level set is applied to Soft multisets. In addition, some theorems are established and proved.

References

Atmaca, S. & Zorlutuna, I. (2013). On fuzzy soft topological spaces, *Annals of Fuzzy Mathematics an Informatics,* 5, 1, 377- 386.

Zou, Y., Xiao, Z. (2008). Data analysis approaches of soft sets under incomplete information, *Knowledge-Based Systems*, 21, 941-945.

Majumdar, P. & Samanta, S.K. (2008). Similarity measure of soft sets, *New math. & natural computation*, 4, 1, 1-12.

Maji, P. K., Roy, A. R. & Biswas, R. (2002). An application of soft sets in a decision making problem, *Computers Math. With Appl.,* l, 44, 1077-1083.

Knuth, D. E. (1973). The Art of computer programming, Vol. 3, (Sorting and Searching), Addison-Wesley, Reading Mass.

Blizard, W. D. (1991). The Development of Multiset Theory, *Modern Logic*, 1, 319-352.

Singh, D., Ibrahim, A. M., Yohanna, T. & Singh, J. N. (2007). An Overview of The Applications of Multisets, *Novi Sad J. Math*, 37, 2, 73-92.

Singh, D. & Isah, A. I. (2015). Some Algebraic Structures of Languages, *Journal of mathematics and computer science* 14, 250-257.

Isah, A. I. & Tella, Y. (2015). The Concept of Multiset Category, *British Journal of Mathematics & Computer Science* 9, 5, 427-437.

Alkhazaleh, S., Salleh, A. R. & Hassan, N. (2011). Soft Multisets Theory, *Applied Mathematical Sciences*, 5, 72, 3561 – 3573.

Majumdar, P. (2012). Soft Multisets, *J. Math. Comput. Sci*. 2, 6, 1700-1711.

Babitha, K. V. & Sunil, J. J. (2013). On soft multi sets, Annals of Fuzzy Mathematics and Informatics, 5, 1, 35-44.

Tokat, D. & Osmanoglu, I. (2013). Connectedness on Soft Multi Topological Spaces, *J. New Results Sci.* 2, 8-18.

Blizard, W. D. (1989). Multiset Theory, *Notre Dame Journal of Formal Logic*, 30, 36-66

S . K. Nazmul, P. Majumdar and S . K. Samanta (2013). On Multisets and Multigroups, Ann. Fuzzy Math. Informa. 6, 3, 643-656.

A. M. Ibrahim, J. A. Awolola, A. J. Alkali (2016). An extension of the concept of n-level sets to multisets, Annals of Fuzzy Mathematics and Informatics, Volume 11, 6, 855-862

Molodtsov, D. (1999). Soft set theory-First results, *Comput. Math. Appl*., 37, 4/5, 19-31.

Sezgin, A. & Atagun, A.O. (2011). On operations of soft sets, *Computers and mathematics with applications,* 60, 1840-1849.

Jena, S. P., Ghosh, S. K. & Tripathy, B.K. (2001). On the theory of bags and lists, *Inform. Sci*. 132, 241-254.

K. P. Girish and S. J. John, General relations between partially ordered multisets and their chains and antichains, Math. Commun. 14(2) (2009) 193-206.