Disease distribution modelling

JELENA LUKIC
Faculty of Technology and Metallurgy, Belgrade University, Karemujeva 4, 11000 Belgrade
SERBIA
jelenapiano@yahoo.com

Abstract: In this paper disease distribution was modelling by randomize method. Cancer distribution in human population was modelling by polynomial distribution probability model. By the derived model virtual experiments of cancer tissue investigation were simulated. The method consists from data collecting about experimental diagnostics, experimental design establishing, selection of mathematical description, parameters estimation, and model adequate examination. How do you seek out sampling function is one of the basic task of modelling this medical system. It is very important how are estimation distribution parameters performed and how defined operation variables according to selected object function. As results of the repeated measurements for the most real experiments different values of measured quantities are obtained.

Keywords: Diagnostic, population, probability, distribution, cancer, sample taking.

1 Introduction
Discrete variables can have only particular values in some intervals and all of them have the same distribution probability function, and they present sample. Each function of their parameters represents sampling function [1]-[4]. Organizations themselves are changing rapidly. Such change is being influenced by a number of outside factors, which makes it difficult to compare current progress with past events. Particular factors which relect these changes include shortening of time horizons - internal and external pressures, reduction and communication.

In the framework of investigation in this paper a model distribution probability cancer for patients different age levels was derived. The derived model adequate describes cancer distribution for different kind of population and geographical origin. Base on the determined parameters from data of experimental investigation it was simulated and predicted values for the other patients groups from the same domain. With the derived models simulations were performed and virtual presentation of the future events at the new patients groups were provided. Simulation of examined phenomena by programme module was performed which made for this purposes.

2 Sampling function definition
Let consider independent variables,

\[ X_1, X_2, X_3, ..., X_n \]

if each of them has the same distribution then they represent sample of the distribution. Each parameters function

\[ Y = f(X_1, X_2, X_3, ..., X_n) \]

represents sampling function. How does choose this sampling function is the main task of disease distribution modeling. It’s very important which method has used for parameters estimation and how to define operation variables according to selected object function. One of them is method for assembly characteristics estimation.

If could be diagnostic process simulated need it’s adequate mathematical description. Here mathematical modeling begin with
statement problem formulation in the form of probability equations system. Mathematical modeling has finished by checking accuracy of the solution.

Modelling method involves the experimental data collecting, experimental design making, mathematical description selecting, parameters estimation in the model and examination of model adequateness by Fisher’s statistical test and checking significance of the coefficients by t-test.

3 Disease diagnostic model

In a case when results of examination can be one of k responds polynomial distribution can be use [4].

Polynomial distribution describes process sampling for k>2. Probability the first sample taking \( x_1 \) is equal \( p_1 \), probability the second sample taking \( x_2 \), is equal \( p_2 \), etc., \( p_1, p_2, p_3, \ldots, p_k \) and,

\[
\sum_{i=1}^{k} p_i = 1
\]

and that means probability of \( n \) examined events, \( A_1 \) \( x_1 \) times, \( A_2 \) \( x_2 \) times, \( A_3 \) \( x_3 \) times, \( A_i \) \( x_i \) times, is equal:

\[
f(x_1, x_2, x_3, \ldots, x_k; n, p_1, p_2, p_3, \ldots, p_k) = \frac{n!}{x_1!x_2!x_3!\ldots x_k!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \ldots p_k^{x_k} \tag{2}
\]

for \( x_1, x_2, x_3, \ldots, x_k = 0,1,2,3,\ldots, n \), and

\[0 \leq p_1, p_2, p_3, \ldots, p_k \leq 1,
\]

and,

\[
\sum_{i=1}^{k} x_i = n
\]

Each experiment need to be independent and probability of each respond need to be uniform, then expression,

\[
\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{3}
\]

Represents number of combination which presents number of different way in which \( x \) responds can happen of total \( n \).

If \( P(x_1, x_2, x_3, \ldots, x_n) \) is probability for \( n \) experiments to be \( x_i \) positive events for variable \( i \), then the probability distribution is equal.

\[
P_n(x_1, x_2, x_3, \ldots, x_n) = \frac{n!}{x_1!x_2!\ldots x_n!} p_1^{x_1} p_2^{x_2} \ldots p_n^{x_n} \tag{4}
\]

with mean value

\[
\mu_x = np_i \tag{5}
\]

and dispersion,

\[
\sigma_x^2 = np_i (1 - p_i) \tag{6}
\]

4 The cancer distribution parameters determination

By derived stochastic model various age groups were investigated from different demographic regions.

As cases studies have using samples from muscle’s tissue, fat tissue, and inner organs for examination to cancer.

4.1 The first case study

Let consider the first example, 12 examined patients age 40-50 years. Experimental investigation show cancer in the muscle’s tissue- event \( A_1 \) found out three times \( x_1=3 \), in the fat tissue-event \( A_2 \) found out four times \( x_2=4 \), and in the inner organs-event \( A_3 \) found out five times \( x_3=5 \).
If it is assuming probability cancer appear in the muscle’s tissue \( p_1 = 0.25 \), probability cancer appear in the fat tissue \( p_2 = 0.33 \), and probability cancer appear in the inner organs \( p_3 = 0.41 \) then distribution function cancer appearing at the noted tissues of the patients is equal:

\[
P_{12}(x_1, x_2, x_3) = \frac{12!}{x_1!x_2!x_3!} \cdot 0.25^{x_1} \cdot 0.33^{x_2} \cdot 0.41^{x_3} \quad (7)
\]

Basic statistical parameters for the first case study can be determined as follows.

Moment, mean values, cancer appearing in the muscle’s tissue is

\[
\mu_{x_1} = np_1 = 12 \cdot 0.25 = 3.0
\]

For the fat tissue,

\[
\mu_{x_2} = np_2 = 12 \cdot 0.33 = 3.96
\]

and for the inner organs

\[
\mu_{x_3} = np_3 = 12 \cdot 0.41 = 4.92
\]

Corresponding dispersion is deriving from equation (6).

\[
\sigma_{x_1}^2 = np_1(1 - p_1) = 12 \cdot 0.25(1 - 0.25) = 2.25
\]

\[
\sigma_{x_2}^2 = np_2(1 - p_2) = 12 \cdot 0.33(1 - 0.33) = 2.65
\]

\[
\sigma_{x_3}^2 = np_3(1 - p_3) = 12 \cdot 0.41(1 - 0.41) = 2.90
\]

With these parameters can make prediction disease expansion in the next period.

4.2 The second case study

The second patient group age 30-40 years, for group of 18 patients, cancer in muscle’s tissue- \( A_1 \) appears \( x_1 = 7 \) times, in fat tissue – \( A_2 \) appears \( x_2 = 9 \) and in inner organs \( A_3 \) appears \( x_3 = 4 \).

If it is assuming probability cancer appear in muscle’s tissue \( p_1 = 0.38 \), probability cancer appear in fat tissue \( p_2 = 0.5 \), and probability cancer appear in inner organs \( p_3 = 0.22 \) then distribution function cancer appearing at the tissues of the patients is equal:

\[
P_{18}(x_1, x_2, x_3) = \frac{18!}{x_1!x_2!x_3!} \cdot 0.38^{x_1} \cdot 0.5^{x_2} \cdot 0.22^{x_3} \quad (8)
\]

Basic parameters for the second case of examination are:

Mean values of cancer appearing or moment, in the muscle’s tissue is

\[
\mu_{x_1} = np_1 = 18 \cdot 0.38 = 6.84
\]

for the fat tissue,

\[
\mu_{x_2} = np_2 = 18 \cdot 0.5 = 9.0
\]

and for the inner organs

\[
\mu_{x_3} = np_3 = 18 \cdot 0.22 = 3.96
\]

Corresponding dispersion is deriving from equation (6).

\[
\sigma_{x_1}^2 = np_1(1 - p_1) = 18 \cdot 0.38(1 - 0.38) = 4.2
\]

\[
\sigma_{x_2}^2 = np_2(1 - p_2) = 18 \cdot 0.5(1 - 0.5) = 4.5
\]

\[
\sigma_{x_3}^2 = np_3(1 - p_3) = 18 \cdot 0.22(1 - 0.22) = 3.08
\]

With these parameters can make prediction disease expansion in the next period.

4.2 The third case study

The third patient group age 20-30 years, for group of 21 patients, cancer in muscle’s tissue- \( A_1 \) appears \( x_1 = 6 \) times, in fat tissue – \( A_2 \) appears \( x_2 = 8 \) and in inner organs \( A_3 \) appears \( x_3 = 7 \).

If it is assuming probability cancer appear in muscle’s tissue \( p_1 = 0.28 \), probability cancer appear in fat tissue \( p_2 = 0.38 \), and probability cancer appear in inner organs \( p_3 = 0.33 \) then distribution function cancer appearing at the tissues of the patients is equal:
Basic parameters for the second case of examination are:

Mean values of cancer appearing or moment, in the muscle’s tissue is

\[ \mu_{x_1} = np_1 = 21 \times 0.28 = 5.88 \]  

for the fat tissue,

\[ \mu_{x_2} = np_2 = 21 \times 0.38 = 4.56 \]  

and for the inner organs

\[ \mu_{x_3} = np_3 = 21 \times 0.33 = 6.93 \]  

Corresponding dispersion is deriving from equation (6).

\[ \sigma^2_{x_1} = np_1(1-p_1) = 21 \times 0.28(1-0.28) = 4.58 \]  
\[ \sigma^2_{x_2} = np_2(1-p_2) = 21 \times 0.38(1-0.38) = 4.94 \]  
\[ \sigma^2_{x_3} = np_3(1-p_3) = 21 \times 0.33(1-0.33) = 4.64 \]  

With these parameters can make prediction disease expansion in the next period for the same demographic region and the same ages.

5 Disease distribution simulation

Simulation has been done by programme's module which have had made for this purpose.

A general technique that does produce efficient algorithms is the method of mixtures. The general distribution can be written

\[ P_n(x) = P_{1n}(x) + P_{2n}(x) + P_{3n}(x) \]  

(10)

The simulation for various patients groups and various \( x_i \) and uniform age periods was provided.

6 Results and discussion

The obtained results have shown in the following Figures 1-8. Model for overall disease diagnosis has shown in Fig. 1.

Fig.2 Shows disease distribution for all patients.
Tissue type: 1-muscle, 2-Fat tissue, 3-Inner organs

Fig.3 Distribution of disease appearing at the first group of patients

Fig.4 Shows frequency of disease appearing vs. tissue type for the first group examined patients.

Fig.4 3-D chart of disease appearing for the first patients group

In Fig.5 has shown frequencies disease appearing vs. tissue type for the second group of patients. Fig.6 shows frequencies disease appearing vs. tissue type for the third group of patients

7 Conclusion

This paper describes disease diagnostic system. The sample taking system was modelling. The polynomial distribution successfully represented acceptance sampling function.

1-muscle, 2-Fat tissue, 3-Inner organs

Fig.7 Frequencies of cancer appearing in the third group of patients
The obtained results in the frame of this investigation show information processing about cancer distribution in population.

Results of this investigation can be applied in the other domain of the medical engineering.

**Acknowledgment.** The author wishes to express her gratitude to the Fund of Serbia for financial support.

**Notation**

- **A**-event
- **k**-number of events
- **f**-function of probability distribution
- **P**-probability distribution
- **p**-probability of event appearing
- **x**-satisfactory result

Green letter

- **μ**-mean value (moment)
- **σ**-standard deviation
- **σ²**-dispersion

**8 References**


