Statistical Modelling of Marketing Processes

ZVONKO SAJFERT,
Technical Faculty "M. Pupin", Zrenjanin, Djure Djakovica b.b.,
University of Novi Sad

CARIŠA BEŠIĆ,
Technical Faculty "M. Pupin", Zrenjanin, Djure Djakovica b.b.,
University of Novi Sad

DEJAN ĐORĐEVIĆ,
Technical Faculty "M. Pupin", Zrenjanin, Djure Djakovica b.b.,
University of Novi Sad

VJEKOSLAV SAJFERT
Technical Faculty "M. Pupin", Zrenjanin, Djure Djakovica b.b.,
University of Novi Sad
http://www.tf.zr.ac.yu/persons.aspx
SERBIA

Abstract: - Treating number of customers and time as random variables and taking into account binary behaving of customers in shopping, the general expression for marketing distribution is formulated. Binary probabilities are averaged by means of beta distribution. On the basis of the general formula and some of its specifications, the ways of marketing advantages are proposed. Among the advantage approaches the two ones are stressed: maximum-likelyhood-methode combined with transition to new random variable and method of composition of distributions. The application of theoretical results in practice is discussed in detail.

Key-Words: - model of marketing process, application of mathematical statistics, distribution formula, random variable

1 Introduction

The role of mathematical statistics methods in different domains of human activities becomes, in nowadays, more important and more necessary. By means of statistical methods can be estimated different risk categories (elemental catastrophes, fires, explosions, business, sports etc.), the behaviour of different business trends can be predicted, the expected result of elections can be determined and so on. The successful application of mathematical statistics methods in mentioned activities is the reason for including of these methods in marketing processes. It can be done in most wide sense starting from selling in shops up to interactions of multinational companies [1,2].

Our goal is to formulate a model of marketing processes, which would satisfy the following conditions:

- distribution formulae of the model have to be relatively simple and convenient for rapid estimate of marketing situation,

- model has to be "elastic", i.e. adoptable to maximal set of different marketing situations,

- the starting formulae of the model have to be accessible rapid changes, capable to follow changes of marketing trends,

- model has to be as realistic as it is possible.

The starting assumption of the model is that basic random variables in marketing are time intervals and number of customers. The behaviour of customers has binary character: the customer buys goods or does not buy it.

On the basis of the model formulated by mentioned requirements and obtained distribution function together with its main characteristics (mathematical expectation and dispersion) some elementary estimates were done. Besides, different possibilities of advancing of model are proposed.

The aim of these advances is better adopting of the model to real situation. Attempts of advancing used, are composition of starting distributions as well as the linear transformation of random variable combined with Maximum-likelyhood-methode.

Finally, an illustrative example of profit estimate as well as the possibilities of bankrupt is quoted.

2. Formulation of the Model

In introductory part it was said that time and number of customers are random variables of the model and that binary behaving of customers is assumed. It is the reason to start the formulation of method from general formula for analogous situation, which most often arises in problems of life-insurance [1]:

$$W_{k+1}(j) = q_{k+1} W_k(j-i) + p_k W_k(i) \tag{1}$$

In this formula W is distribution density of life-insurance persons, index j counts time units, index k counts insuranced persons, p is surviving probability of insuranced persons in some time period, while q is dying probability. It is clear that $p+q=1$ and that p and q are dependent on k as well as on additional time index which is denoted with i.

The quoted formula (1) is very complicated for calculations. In order to simplify formula (1) we shall substitute p_k and q_k with some constant average values p and q. The problem of averaging p and q will be discussed later. In accordance with this (1) goes over to:

$$W_{k+1}^{i}(j) = qW_k^{j-i,i}(j-i) + pW_k^{i,i}(i) \tag{2}$$

Formula (2) is double recurrent. For fixed $j = j_0$ and $i=0$, and initial condition $W_0(0)=0$, it reduces to $W_{k+1}(j_0) = qW_k(j_0)$, wherefrom it follows $W_k(j_0) = q^k W_0(j_0)$. If we take $i = j_0$ formula (2) reduces to $W_k(j_0) = p^k W_0(j_0)$. Obtained expressions are valid for all values of k. If $i \in [1, j_0 - 1]$ formula (2) gives combinations $W_k(j_0) = qW_k(j_0 - i) + pW_k(i)$.

Further formulation of the model is based on the assumption that the probabilities $W_k(j_0 - i)$ and $W_k(i)$ are proportional to magnitude of time intervals: $W_k(j_0 - i) \sim j_0 - i$ and $W_k(i) \sim i$. This assumption is quite logical and, by the way, such assumption was the basic one for deriving of Poisson's distribution. Taking into account that every W_k is proportional to q^k or p^k, we choose proportionality factor as arithmetic average of p^k and q^k, i.e. $\frac{1}{2}(p^k + q^k)$. This leads to the formula, valid for arbitrary j,

$$W_k^{(i)}(j) = \frac{1}{2}(p^k + q^k)[q(j-i) + pi] \tag{3}$$

The last stage of the formulation of the model is the following: distribution density that k customers take part in shopping in time moment j is given as a sum distribution densities $W_k^{(i)}(j)$ over all values of time $i = 0,1,2,..., j$. This distribution density will be denoted with $G_k(j)$. So we obtain:

$$G_k(j) = \frac{1}{4}(p^k + q^k) j(j+1).$$

Distribution density G has to be normalized. If we take that marketing process is lasting for J time units and that N (N is of the order 10^3 - 10^4) is total number of customers, we can write:

$$C\sum_{j=0}^{J} \sum_{k=0}^{N} \frac{1}{4}(p^k + q^k) j(j+1) = 1. \tag{4}$$

Calculating the sums we find

$$\sum_{j=0}^{J} j(j+1) = \frac{1}{3} J(J+1)(J+2) \tag{5}$$

and

$$\sum_{k=0}^{N} \left(p^k + q^k\right) = 1 - \frac{1-p^{N+1}}{1-p} + \frac{1-q^{N+1}}{1-q} \approx \cdot$$

$$\approx 1 - \frac{1}{1-p} + \frac{1}{1-q} = \frac{1}{pq} \tag{6}$$

This gives:

$$C = \frac{12pq}{J(J+1)(J+2)} \tag{7}$$

Consequently, the normalized distribution density G is given by

$$G_k(j) = \frac{3pq}{J(J+1)(J+2)} j(j+1)(p^k + q^k) \tag{8}$$

This formula represents the basic one of our model and it defines probability density to k customers participate marketing in time moment j.

Although the procedure of obtaining the

model distribution *G* was exposed in detail, we wish to point out that the model distribution is not the solution of formula (1). Formula (1) only inspired us to derive the expression for model distribution *G*. In formula (1) was made a wide set of approximations. Besides, it was assumed that distribution density is proportional to the magnitude of time interval. Simplified formula (1) was used for determining of proportionality factor form and this is maximum of formula (1) participation in the model formula (8).

It is of interest for further analyses to determine, separately, time-dependent probability density Ψ_j and probability density depending on number of customers *k* only Φ_k.

According to the rules of probability theory, the probability Ψ_j represents the sum of $G_k(j)$ over all values of *k*. This summing is very simple and it will not be quoted. We quote only the final result:

$$\Psi_j = \frac{3j(j+1)}{J(J+1)(J+2)} \tag{9}$$

Summing *G* over all values of *j* we get distribution density depending on number of customers

$$\Phi_k = pq\left(p^k + q^k\right) \tag{10}$$

Application of formulas (8), (9) and (10) will be analysed later. Here we shall pay attention to determining average values of binomial parameters *p* and *q*.

There exist a series of statistical distributions, which could be used for averaging parameters *p* and *q*, but beta distribution seems to be the most convenient due to two reasons. One of reasons is the fact that beta distribution [1] is two-parameters one and this gives wider possibility for more real choice of averages. Second, more important reason, is connected with the fact that random variable *x* lies in the interval $[0,1]$. The binomial parameters *p* and *q* lie in this interval, also.

The density of beta distribution is given by:

$$B_{\alpha\beta}(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}; \quad x \in [0,1] \tag{11}$$

In this formula α and β are beta distribution parameters, while Γ is gamma function [3,4]. The expected value random variable *X* which will be denoted with $\hat{E}X$, where $\hat{E}$ is the operator $\int_0^1 dx B_{\alpha\beta}(x)(\ldots)$, is: $\hat{E}X = \frac{\alpha}{\alpha+\beta}$.

Beta distribution has one maximum and its location is defined by the condition:

$$(\alpha-1)(1-x_m) = (\beta-1)x_m \tag{12}$$

which is obtained by equating $\frac{\partial B}{\partial x}$ with zero.

Here will be quoted a set of illustrative examples for defining average values of *p* and *q*.

A year has twelve months so that for time unit will be taken one month. It is known that sale is increased in September (ninth month) because customers are buying vegetables and fruit for winter period. It is reason to determine beta distribution parameters α and β for $x_m = \frac{9}{12} = \frac{3}{4}$. We shall take that $p = x$ and $q = 1 - x$. Since *p* has to be higher than *q* (due to increased sale in September) we shall take $\beta = 2$. For this value β, on the basis of (12), we obtain $\alpha = 4$ and in such a way we finally have $p = \frac{2}{3}$ and $q = \frac{1}{3}$.

Shortly before New Year Day (we shall take $x_m = \frac{29}{30}$ and $\beta = 2$), the sale is very intensive. For quoted x_m and β we obtain $\alpha = 30$. Including this into expression for expectable values we get $p = \frac{15}{16}$ and $q = \frac{1}{16}$.

Now we shall consider an example corresponding to the minimum sale. It is the period between 1. January and 10. January, usually. For this case we can take $x_m = \frac{1}{20}$ and $\alpha = 2$. As earlier we take $p = x$ and $q = 1 - x$. Expectable value of *p* is $\frac{1}{11}$, while for *q* is $\frac{10}{11}$. In similar way can be shown that for $x_m = \frac{1}{4}$ (March) average of *p* is $\frac{1}{3}$, while average of *q* is $\frac{2}{3}$. For $x_m = \frac{1}{2}$ averages of *p* and *q* are equal, i.e. $p = q = \frac{1}{2}$.

The analyses exposed, clearly point out that in second half of year decreasing of the sale decreases *p* and it is expectable result. For minimal sale in the first half of year we obtain *q*>>*p* and this is expectable, also. If we have maximum of sale in first half of year or minimum of sale in second half of year then is necessary to *p* and *q* change roles, i.e.

for intensive sale in first half of year q has to be sale probability and vice verse, for minimum of sale in second half of year p has to be probability of sale cancellation.

The determining of averages p and q must not be done by means of beta distribution. It can be done, in principle, by means of arbitrary distribution having extremum, but the calculations are more complicated.

It is important to note that illustrative examples quoted have practical significance. Using these examples salesman can predict what sale can expect in different time periods and consequently to fight for the optimal regime of selling.

The analysed marketing process lasting one year. The marketing process lasting one day (twelve hours) is very similar so that quoted can be translated to one day marketing process.

It is very important to point out that averaging by means of beta distribution can inspire maximum elastic behaving of salesman with respect to expectable changes of marketing trend in period of intensive sale. He can advantage the sale in mentioned period improving service and quality of goods. The intensifying advertising is necessary, also.

Now we shall calculate statistical some characteristic of distribution Ψ_j and Φ_k.

Statistical distributions are determined by set of characteristics [2, page 18]. We shall not quote all these characteristics since it requires many calculations and much space. Here will be found only two characteristics, which are the most important for marketing processes. Those are mathematical expectation, i.e. average value of random variable and dispersion, which represent measure of deviation with respect to average value. Following the definition of mathematical expectation we can write for the distribution Ψ_j:

$$\hat{E}j = \sum_{j=0}^{J} j\Psi_j = \frac{3}{J(J+1)(J+2)} \sum_{j=1}^{J} \left(j^3 + j^2\right) = \frac{3J+1}{4} \tag{13}$$

On the basis of dispersion definition we have:

$$\hat{D}j = \hat{E}j^2 - \left(\hat{E}j\right)^2 = \frac{3J^3 + 57J^2 - 27J + 47}{80(J+2)} \tag{14}$$

We note that the sums of the type $\sum_{s=1}^{N} s^m$, leading to the results (13) and (14), are taken form [3].

In deriving mathematical expectation and dispersion for distribution density Φ_k we have taken upper sum boundary infinite since number N of customers is very big. We used auxiliary formula

$$S_0 = \sum_{k=0}^{\infty} a^k = \frac{1}{1-a} \ ; \ a < 1 . \tag{15}$$

Two derivations of (15) with respect to a give

$$S_1 = \sum_{k=0}^{\infty} k a^k = \frac{a}{(1-a)^2} \tag{16}$$

and

$$S_2 = \sum_{k=0}^{\infty} k^2 a^k = \frac{2a^2}{(1-a)^3} + \frac{a}{(1-a)^2} . \tag{17}$$

Using (16) and (17) we obtained mathematical expectation in the form

$$\hat{E}k = \sum_{k=0}^{\infty} k\Phi_k = \frac{1}{pq} - 3 \tag{18}$$

and the dispersion as:

$$\hat{D}k = \hat{E}k^2 - \left(\hat{E}k\right)^2 = \frac{p^4(p+1) + q^4(q+1)}{p^2q^2} - \left(\frac{1}{pq} - 3\right)^2 \tag{19}$$

It is interesting to note that $\hat{E}k$ and $\hat{D}k$ have minimal values in the case when $p = q = \frac{1}{2}$. These minimal values are $\hat{E}k = 1$ and $\hat{D}k = 2$.

The analyses of this section will be finished determining the distribution Φ_k in continual approximation. This continual distribution will be denoted with $\Phi(k)$, where k is continual variable. This translation to continuum is justified by very big number of potential customers. The transition symbolically can be expressed as $\Phi_k \to \Phi(k) = \text{const}\left(p^k + q^k\right)$, where, in the last expression, k represents continual variable. The sums have to be substituted by integrals taken form zero to infinity.

The first stage requires normalization of density $\Phi(k)$, i.e.:

$$\int_0^\infty dx\Phi(k) = \text{const}\int_0^\infty dk\left(p^k + q^k\right) = 1\,. \tag{20}$$

Using the formula $a^x = \mathrm{e}^{\ln a^x} = \mathrm{e}^{x\ln a}$ we can write (20) in the form:

$$\text{const}\int_0^\infty dk\left(\mathrm{e}^{-Pk} + \mathrm{e}^{-Qk}\right) = 1\,, \tag{21}$$

where

$$P = -\ln p \in (0,\infty); \tag{22}$$

$$Q = -\ln q \in (0,\infty).$$

Integrals in (21) can be easily solved. Solving integrals we get $\text{const} = \dfrac{PQ}{P+Q}$. Consequently for normalized distribution density we obtain:

$$\Phi(k) = \frac{PQ}{P+Q}\left(\mathrm{e}^{-Pk} + \mathrm{e}^{-Qk}\right); \tag{23}$$

$$P = -\ln p\,;\; Q = -\ln q\,.$$

Using this continual variant of density we shall determine average value of consumers number $\hat{E}k$ and the deviation of this average value $\hat{D}k = \hat{E}k^2 - \left(\hat{E}k\right)^2$. These averages will be found by means of generating function. The generating function for continual distribution density $f(x)$ is defined as

$$\gamma(t) = \int_a^b dx e^{tx} f(x). \tag{24}$$

The importance of generating function consists in the fact that statistical moments of the order n, i.e. $\hat{E}x^n$ are determined by n-th derivative of this function taken for $t = 0$, i.e.:

$$\hat{E}x^n = \frac{d^n\gamma}{dt^n}\Big|_{t=0}\,. \tag{25}$$

Generating function for distribution $\Phi(k)$ is given by

$$\gamma_\Phi(k) = \int_0^\infty dk e^{tk}\Phi(k) =$$

$$= \frac{PQ}{P+Q}\int_0^\infty dk\left[e^{-(P-t)k} + e^{-(Q-t)k}\right] =$$

$$= \frac{PQ}{P+Q}\left(\frac{1}{P-t} + \frac{1}{Q-t}\right). \tag{26}$$

Since we are looking for mathematical expectation and dispersion we shall by means of the formulas (25) and (26) determine the first order moment $\hat{E}k$ and the second order moment $\hat{E}k^2$. Since

$$\frac{d\gamma_\Phi}{dt} = \frac{PQ}{P+Q}\left[\frac{1}{(P-t)^2} + \frac{1}{(Q-t)^2}\right]. \tag{27}$$

and

$$\frac{d^2\gamma_\Phi}{dt^2} = 2\frac{PQ}{P+Q}\left[\frac{1}{(P-t)^3} + \frac{1}{(Q-t)^3}\right]. \tag{28}$$

we immediately obtain

$$\hat{E}k = \frac{d\gamma_\Phi}{dt}\Big|_{t=0} = \frac{P^2+Q^2}{PQ(P+Q)} \tag{29}$$

and

$$\hat{E}k^2 = \frac{d^2\gamma_\Phi}{dt^2}\Big|_{t=0} = 2\frac{P^3+Q^3}{P^2Q^2(P+Q)} \tag{30}$$

Using (29) and (30) we obtain the expression for dispersion

$$\hat{D}k = \hat{E}k^2 - \left(\hat{E}k\right)^2 =$$

$$= \frac{1}{P^2Q^2(P+Q)}\left[2P^3 + 2Q^3 - \frac{\left(P^2+Q^2\right)^2}{P+Q}\right] \tag{31}$$

The determined averages (29) and (31) are expressed in terms of parameters P and Q. They have to be expressed in terms of parameters p and q. It can be easily achieved by substitutions $P = -\ln p$ and $Q = -\ln q$. Since this substituting is simple we shall not quote the expressions (29) and (31) in terms of p and q.

In future analyses the attention will be paid to some methods with could be able to advance distributions Ψ_j, Φ_k and $\Phi(k)$. In order to achieve that, we shall use the method of composition (convolution) of distributions and method of going over to new random variables combined with Maximum-likelyhood-methode.

4 Conclusion

The analysis of application of the mathematical statistics methods in marketing enveloped a set of approaches, which could advance the marketing process. If reader asks: what are the conditions of successful marketing, our answer is the following: these conditions are known a long time ago. Those are high quality of goods, high quality of serve, lowering of prices, advertisement of goods, the permanent following of marketing trends and general financial situation etc. Since the quoted conditions are known, the following question is natural: What is the benefit of such abundant analysis? The answer is clear: the analysis gives series of instructions when and how must be acted in the marketing process advantage direction.

Acknowledgement

This work was supported by the Serbian Ministry of Science and Technology: Grant No. 141044.

References:

[1] Hipp C. (1994). *Introduction to risk theory*, Lectures on Technical University Karlsruhe,

[2] Bühlmann H. (1970). *Mathematical Methods in Risk Theory*, Springer-Verlag, Berlin, Heidelberg.

[3] Grashteyn I.S. , Ryzhik I.M. (1965). *Table of Integrals, Series, and Products*, Academic Press, New York, London.

[4] Korn G. ,Korn T. (1968). *Mathematical Handbook for Scientists*, Dover Publ.

[5] Krishnamoorthy A. , Misra S. , Prasad A. (2005). Scheduling sales force training: Theory and evidence, *International Journal of Research in Marketing*, **22**, 427-440

[6] Beard R. , Pentikonen T. , Pesonen E. (1969). *Risk Theory*: *The Stochastic Basis of Insurance*, Metheun, London

[7] Abramowitz M. (1965). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Table, Dover