# Application Bayesian Approach in Tasks Decision-Making Support

**Aliyeva Tarana Abulfaz***

Department of Information Economic and Technology, Azerbaijan State Economic University, Baku, Azerbaijan
*Corresponding author: alilitarana@gmail.com

**Abstract** The proposed work is devoted to the use of the Bayesian approach based on probabilistic method of use, along with the original statistical data prior information about the current process. Especially significant advantages over classical methods in terms of the accuracy of the statistical inference Bayesian approach is in the relatively small samples, which is very characteristic of the simulation. A theorem on minimizing classification error probability is proved. About choosing a priori distribution parameters from the coordinated class of distributions is considered.

*Keywords:* a priori distribution, a posteriori distribution, a coordinated distribution, Bayesian approach

## 1. Introduction

Many statistical problems independently of the methods of solving them have a common property: before you get a specific set of data as a potentially acceptable to study the situation examines the several probabilistic models. When the data is received, there is a pronounced in a form of knowledge about relative acceptability of these models. One way to "revision" on the relative acceptability of probabilistic models is the Bayesian approach, which is based on Bayes' theorem.

The Bayesian approach has a number of advantages that make it attractive enough for widespread application. The main difference of this approach from other approaches is that before the data is received, the decision-maker or a statistician examines the degree of his confidence in the possible models and presents them in the form of probabilities.

Once the data is received, the Bayes' theorem allows us to calculate a new set of probabilities, which are revised the degree of credibility of possible models taking into account the new information received through the data.

As is known, the statistical datas are often absent in actual tasks Decision-Making support, which makes use of many traditional frequency approaches unlawful. Available information may only contain subjective assessments in the form of expert evaluations and judgments. Moreover, the situation in which it is decided, may be generally new at all and has never been previously analyzed. These characteristics complicate the tasks Decision-Making support process and may call into question any conclusions and opinions. In such a situation, the application of the Bayesian Approach [1] is very effective.

## 2. Experimental Methods

The Bayesian approach [2] is based on the statistical nature of the observations. It is known that in the decision statistic used as input simultaneously two types of information: a priori (before the existing statistical production) and is contained in the initial statistical data. In this case, a priori information given to him in the form of an a priori probability distribution of the analyzed unknown parameter $\theta$, that describes the degree of confidence that this option will take some value, before the collection of baseline statistics. State the value of the unknown parameter of the system means to find its posteriori distribution. The idea of the Bayesian approach lies precisely in the transition from a priori distribution to the posterior distribution.

Let $A_1, A_2, ..., A_n$ - full group of mutually exclusive events. $\bigcup_{i=1}^{n} A_i = \Omega$, $A_i \cap A_j = \varnothing$, when $i \neq j$. Then the posterior probability is:

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)}, \qquad (1)$$

where $P(A_i)$ - the a priori probability of the event $A_i$, $P(B \mid A_i)$ - the conditional probability of an event $B$, provided that the event occurred $A_i$, and the event $B$ has a non-zero probability ( $P(B) > 0$ ).

Initially, Bayesian classification is used to formalize the knowledge of experts in expert systems, Bayesian classification is now also used as a method of Data Mining.

The Bayesian approach consists of the following steps:

1. Determination of the a priori distribution $p(\theta)$ of the desired multi-dimensional parameter $\theta$;

2. Getting the source of statistical data $x_1, x_2, ...., x_n$ to the laws of distribution $f(x_i|\theta), i = 1, 2, ..., n$ at fixed parameter $\theta$;

3. The calculation of the likelihood function $L(x_1, x_2, ...., x_n|\theta)$ defined by the relation

$$L(x_1, x_2, ...., x_n|\theta) = f(x_1|\theta)f(x_2|\theta)......f(x_n|\theta) \quad (2)$$

4. The calculation of the posteriori distribution $\tilde{p}(\theta|x_1, x_2, ...., x_n)$, determined by the relation

$$\tilde{p}(\theta|x_1, x_2, ...., x_n) = \frac{p(\theta)L(x_1, x_2, ...., x_n|\theta)}{\int L(x_1, x_2, ...., x_n|\theta)p(\theta)d\theta} \quad (3)$$

5. Building a Bayesian point and interval estimates of parameters $\hat{\theta}$ using the mean or modal value of the distribution found from the formula (3).

$$\hat{\theta}_{mean} = E(\theta|x_1, x_2, ...., x_n) = \int \theta \tilde{p}(\theta|x_1, x_2, ...., x_n)d\theta \quad (4)$$

$$\hat{\theta}_{mod} = \arg\max_{\theta} \tilde{p}(\theta|x_1, x_2, ...., x_n).$$

## 2.1. Experimental Set Up

In many decision-making tasks a priori probability information about the States of nature can be changed after the new expert assessments or as a result of observation relevant developments related to states and confirming or refuting a priori information.

The dependence of the posterior probabilities of a priori shows how much information and values of the unknown parameters contained in the statistics. If the posterior probabilities are highly dependent on the a priori, it is likely the data contains little information. If the posterior probabilities are weakly dependent on the choice of a priori distribution, the data are informative.

Thus, using the Bayesian approach except for the probability distribution of the considered random variable $X$ is assumed that some a priori distribution parameters $\theta$, the distribution function of values $X$. Relying on statistical data, the a priori distribution of the parameters $\theta$ is modified by multiplying the likelihood function and normalization. The result of the modification is the posterior distribution of the parameters $\theta$. In other words, the parameters of the random variable themselves are random variables with some distribution.

The most important and challenging at the same time is a matter of choosing a priori distribution parameters. One of the factors here is that the presence of even statistical informative "poor" a priori distribution will not significantly affect the posterior. Another important factor is the computational complexity, especially if the calculations of the posterior distribution produced in series as they become available statistical information. Therefore, the choice of the a priori distribution affects its belonging to so-called class coordinated distributions such distributions that are a priori and a posteriori the same distribution, but with different parameters.

Priori distributions modeling no a priori information are called uninformative. The postulate Bayes-Laplace [3]

says that when nothing is known about the pre-parameter $\theta$ a priori distribution should be uniform, i.e., all manner of outcomes of a random variable $\theta$ have equal probability. The main problem of the use of uniform distribution as a non-informative a priori distribution is that a uniform distribution is not invariant with respect to the function of the parameter. If we know nothing about the parameter $\theta$, then we also do not know, for example, about the function $1/\theta$. However, if $\theta$ has a uniform distribution, $1/\theta$ no longer has a uniform distribution, although according to the postulate Bayes - Laplace, $1/\theta$ must have the uniform distribution. Moreover, the uniform distribution cannot be used as a priori, if the set parameter values are infinite.

In the literature, there is enough considerable number of approaches for the selection of the no information priori distribution having its advantages and disadvantages. However, the most interesting is, the approach is completely different from most traditional. The essence of this approach is as follows. Define more than one prior distribution, but a whole class $\mathfrak{Z}$ of distributions $\pi$, for which you can find lower and upper probability of the event $A$ as a

$$\underline{P}(A) = \inf\{P_\pi(A) : \pi \in \mathfrak{Z}\}$$

$$\overline{P}(A) = \sup\{P_\pi(A) : \pi \in \mathfrak{Z}\}.$$

Under certain conditions, the set is completely determined by the lower and upper probability distribution functions. It should be noted that the class should be seen "not as a class of suitable prior distributions, as well as a suitable class of prior distributions. This means that every single distribution of the class is not the best or "good" a priori distribution, since no single distribution can not satisfactorily simulate the absence of information. But the whole class as a whole, define the upper and the lower the probability distribution is an appropriate model for the lack of information.

**Definition.** The family of a priori distributions $G = \{p(\theta; D)\}$ is called the conjugate with respect to the likelihood function $L(x_1, x_2, ...., x_n|\theta)$, and if the posterior distribution $\tilde{p}(\theta|x_1, x_2, ...., x_n)$ as calculated by the formula (3), again belongs to the same family $G$.

If the likelihood function $L(x_1, x_2, ...., x_n|\theta)$ can be represented in a form

$$\begin{aligned} &L(x_1, x_2, ...., x_n|\theta) \\ &= v\begin{pmatrix} T_1(x_1, x_2, ..., x_n), ...., \\ T_m(x_1, x_2, ..., x_n); \theta \end{pmatrix} \cdot \psi(x_1, x_2, ..., x_n), \end{aligned} \quad (5)$$

Where $T_1(x_1, x_2, ..., x_n)$, $j = \overline{1, m}$ and $\psi(x_1, x_2, ..., x_n)$ - some of the functions of the observations $x_1, x_2, ..., x_n$, do not depend on the parameters $\theta$, then there is the family $G = \{p(\theta; D)\}$ a priori distributions conjugate to $L(x_1, x_2, ...., x_n|\theta)$. The functions $T_1(x_1, x_2, ..., x_n)$, $j = \overline{1, m}$ also called sufficient statistics.

On the other hand, before any data or observations of parameters are often not known, i.e. the researcher does

not have any useful a priori information about the values of the estimated parameter. In such cases it is necessary to consider the following selection rules [4] priori distribution:

If the estimated scalar parameter $\theta$ can take values in a finite interval $[\theta_{\min};\theta_{\max}]$ or a infinite interval from $-\infty$ to $\infty$, then the a priori density function $p(\theta)$ should be considered constant at the appropriate interval;

if the meaning of the estimated parameter follows that it can take any positive values, it should be considered constant on the real line density function distribution of the logarithm of the value of the parameter, i. e, $p(\ln\theta) = const$, when $\theta \in (0;\infty)$.

This a priori distribution is denoted by $\hat{p}(\theta)$. The fact that so defined on infinite direct or semi-direct a priori distributions violated known rules of normalization of the probability density function (because the $\int \hat{p}(\theta)d\theta \neq 1$, $\int \hat{p}(\theta)d\theta = \infty$, where the integration is over all possible values), does not deliver "technical inconveniences". As a priori distributions used uniform distribution $\hat{p}(\ln\theta) = const$ and it allows the density distribution function $\hat{p}(\theta)$ to minimize the entropy measure of information - $H = \int\limits_{-a}^{a} p(\theta)\ln p(\theta)d\theta$ for arbitrarily large values $a$.

Given the normalization of the density function (3) we obtain:

$$\tilde{p}(\theta|x_1,x_2,....,x_n) \sim p(\theta)\cdot L(x_1,x_2,....,x_n|\theta). \quad (6)$$

Us determine the form a priori density $p(\theta)$ for the case $p(\ln\theta) = const$, when $\theta \in (0;\infty)$.

Let $F_\theta(y) = P\{\theta < y\}$ - distribution function of the parameter. Then

$$F_\theta(y) = P\{\theta < y\} = P\{\ln\theta < \ln y\} = F_{\ln\theta}(\ln y)$$

Accordingly, the density function of the parameter $\theta$ is:

$$f_\theta(y) = \frac{\partial F_\theta(y)}{\partial y} = \frac{\partial F_{\ln\theta}(\ln y)}{\partial(\ln y)} \cdot \frac{\partial(\ln y)}{\partial y}$$
$$= f_{\ln\theta}(\ln y)\cdot\frac{1}{y} \sim \frac{1}{y},$$

since by the condition $f_{\ln\theta}(\ln y) = p(\ln\theta) = const$. So that for $\theta \in (0;\infty)$ we have:

$$\hat{p}(\theta) \sim \frac{1}{\theta}. \quad (7)$$

From this we obtain formula for the calculation of the posterior distribution

$$\hat{p}(\theta|x_1,x_2,....,x_n) \sim \hat{p}(\theta)\cdot L(x_1,x_2,....,x_n|\theta), \quad (8)$$

where the likelihood function $L(x_1,x_2,....,x_n|\theta)$ has the form (5).

**Remark.** Use as a priori laws of probability distributions associated with the observed general set of a family of sets of priori distributions $\{p(\theta;D)\}$. However, the implementation of Bayesian approach is necessary to operate the specific priori distribution, which requires knowledge of the numerical values $D_0$ of the parameters $D$ on which our a priori probability distribution depends. In a broad sense, the a priori distribution parameters can be determined by the method of moments with known average values of the estimated parameters $\theta_0 = E\theta = (E\theta_1,E\theta_2,...,E\theta_s)^T$ and their standard deviations $\Delta_1 = \sqrt{D\theta_1}$, $\Delta_2 = \sqrt{D\theta_2}$, ..., $\Delta_s = \sqrt{D\theta_s}$ ...,.

Let $\{p(\theta;D)\}$ the family of priori distributions, conjugated with the likelihood function $L(x_1,x_2,...,x_n|\theta)$ of observations available to us and let $D_0$ set parameter values $D$ for the analyzed case. Then, using a series of identical transformations the right side of the relation

$$\tilde{p}(\theta|x_1,x_2,....,x_n) \sim p(\theta;D_0)\cdot L(x_1,x_2,....,x_n|\theta) \quad (9)$$

is given to factors independent of $\theta$ to the species $p(\theta;D(x_1,....,x_n))$, where the last function belongs to the family $p(\theta;D)$, and each of the components $d_j(x_1,...,x_n)(j=1,2,....,q)$ of the vector of parameters $D(x_1,....,x_n)$ is a function of $D_0$ and $\{x_1,....,x_n\}$.

Consistency of distributions is determined not only by the views of a priori distribution, but also the view of the likelihood function, i.e., type of distribution must be maintained when multiplying a priori distribution on the likelihood function with the normalization. Dirichlet distribution also applies to distributions.

For a description of the Dirichlet distribution, consider the standard polynomial model.

Let $\Omega = \{\omega_1,...,\omega_m\}$ the set of possible outcomes, and there is a set of $N$ observations independently selected from $\Omega$ the same probability of each outcome $P\{\omega_j\} = \theta_j$ for all $j = 1,...,m$, where $\theta_j \geq 0$ and $\sum\limits_{j=1}^{m}\theta_j = 1$. The probability that the outcome $\omega_j$ of the $N$ observations will occur $N_j$ time, determined from the well-known formula of polynomial distribution with parameters $\theta_1,...,\theta_m$. However, the parameters $\theta_1,...,\theta_m$ themselves may be random variables and have a distribution or probability density $p(\theta)$. One of the most interesting distributions of the parameters $\theta_1,...,\theta_m$ is a Dirichlet distribution, which is coordinated with a polynomial distribution in the sense that a priori and a posteriori distributions are Dirichlet distributions.

A priori Dirichlet distribution $(s,t)$ for a random vector of probabilities $\theta = (\theta_1,...,\theta_m)$, where $t = (t_1,...,t_m)$, has the probability density function of the probability density of the proposed de Groot (De Groot) [5]

$$p(\theta) = \Gamma(s)\left(\prod_{j=1}^{m}\Gamma(st_j)\right)^{-1}\cdot\prod_{j=1}^{m}\theta_j^{st_j-1}.$$

Here the parameter is the average value (expectation) the probability $\theta_i$; the parameter $s > 0$ determines the

effect of a priori distribution on the posterior probabilities; the vector $t$ belongs to the inner area unit simplex of dimension $m$, which we denote $S(1,m)$; $\Gamma(\cdot)$ - the gamma function, satisfying conditions $\Gamma(x+1) = x\Gamma(x)$ and $\Gamma(1) = 1$. Necessary to note that the variables of the Dirichlet distribution are probabilities $\theta_1,...,\theta_m$, satisfying the condition $\theta \in S(1,m)$. This means that not only the random events are considered, but their probabilities.

After receiving the observation vector $n = (n_1,...,n_m)$, where $n_j$ - the number of observations outcome $\omega_j$, multiplying a priori density on the likelihood function $L(n,\theta)$, we obtain the posterior density:

$$\tilde{p}(\theta|n) \sim p(\theta) \cdot L(n|\theta) = \prod_{j=1}^{m} \theta_j^{n_j + st_j - 1}$$

which can be considered as the density of the Dirichlet distribution $(N + s, t^*)$, where

$$t^* = (t_1^*,...,t_m^*), t_j^* = \frac{n_j + st_j}{N + s}.$$

In other words, the Dirichlet distribution belongs to the coordinated class of distributions and under the recalculation a priori parameters $t$ are converted into $t^*$. This is a very important feature, through which the Dirichlet distribution is widespread in the Bayesian analysis.

## 3. Results and Discussion

### 3.1. Results

*Example 1.*

a) A random variable with unknown parameter $\theta$ value has an exponential distribution. Check the conditions of existence of the conjugate priori distribution, to determine its accessories thereof Class of distributions and the specific values of the parameters of the distribution.

It is known that

$$f(x|\theta) = \begin{cases} \theta e^{-\theta x} & for \quad x \geq 0, \\ 0 & for \quad x < 0. \end{cases}$$

From equation (2) have:

$$L(x_1,x_2,....,x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \theta^n e^{-\left(\sum_{i=1}^{n} x_i\right)\theta}. \quad (10)$$

Therefore,

$$L(x_1,x_2,....,x_n|\theta) = \begin{cases} \theta^n e^{-\left(\sum_{i=1}^{n} x_i\right)\theta} & for \quad x_i \geq 0, \\ 0 & for \quad x_i < 0. \end{cases}$$

Here $m = 1$; $T(x_1,x_2,...,x_n) = \sum_{i=1}^{n} x_i$ - sufficient statistic that confirms the existence of a priori, conjugated with $L(x_1,x_2,....,x_n|\theta)$ the distribution of the parameter $\theta$.

Now will update the belonging of this distribution to a certain class of distributions.

Taking into account (7), from (8) we obtain:

$$\hat{p}(\theta|x_1,x_2,....,x_n) \sim \hat{p}(\theta) \cdot L(x_1,x_2,....,x_n|\theta)$$
$$\sim \theta^{n-1} \cdot e^{-\left(\sum_{i=1}^{n} x_i\right)\theta} \quad (11)$$

The right side of relation (11) states that the family of conjugate priori distributions parameter $\theta$ exponentially distributed the general total belongs to the class of gamma distribution with the density

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \cdot e^{-\beta\theta}, \theta > 0$$

and distribution of parameters $\alpha = n$ and $\beta = \sum_{i=1}^{n} x_i$, respectively.

It is known that the mean value ($E\theta$) and variance ($D\theta$) of gamma - distributions are expressed by the parameters of the distribution $\alpha$ and $\beta$ by the formulas:

$$E\theta = \frac{\alpha}{\beta}, D\theta = \frac{\alpha}{\beta^2}.$$

Substituting into these relations instead $E\theta$ and $D\theta$ respectively given values $\theta_0$ and $\Delta^2$, we obtain as the solutions of a system from the two equations (and relative to $\alpha$ and $\beta$):

$$\alpha = \frac{\theta_0^2}{\Delta^2}, \beta = \frac{\theta_0}{\Delta^2}.$$

In the example above, we define a posteriori distribution, taking into account (10) and the density of gamma distribution:

$$\tilde{p}(\theta|x_1,x_2,....,x_n) \sim \theta^{\alpha-1} e^{-\beta\theta} \cdot \theta^{n-1} \cdot e^{-\left(\sum_{i=1}^{n} x_i\right)\theta}.$$

From here, it is seen that the a posteriori distribution of the parameter $\theta$ again obeys the law gamma - distribution, but with the following parameters:

$$\tilde{\alpha} = \alpha + n, \tilde{\beta} = \beta + \sum_{i=1}^{n} x_i.$$

b) The random variable (the number of successes in $s$ series of Bernoulli trials) $\xi_\theta(s)$ has a binomial distribution, where $\theta$ - unknown probability of success in one such test, and $s$ - the total number of tests in this series of Bernoulli. Check conditions for the existence of the conjugate prior distribution; determine its belonging in the class of distributions and specific parameters of the distribution.

It is known that

$$f(x|\theta) = P\{\xi_\theta(s) = x|\theta\} = C_s^x \theta^x (1-\theta)^{s-x},$$
$$x = 0,1,2,....,s.$$

There are observed such series. According to the formula (2) we have:

$$L\left(x_1, x_2,....., x_n \mid \theta\right) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

$$= \prod_{i=1}^{n} C_s^{x_i} \theta^{x_i} (1-\theta)^{s-x_i} \quad (12)$$

$$= \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{ns-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} C_s^{x_i},$$

where $x_i$ the number of successes in $i$-th series. There

$m = 1$; $T(x_1, x_2,..., x_n) = \sum_{i=1}^{n} x_i$ - sufficient statistic that confirms the existence of a priori, coupled with $L(x_1, x_2,....., x_n \mid \theta)$ the distribution parameter $\theta$. Now, we will specify the distribution of belonging to a particular class of distributions.

Define $\hat{p}(\theta) = 1$ for $\theta \in (0;1)$, then

$$\hat{p}(\theta \mid x_1, x_2,....., x_n) \sim \hat{p}(\theta) \cdot L(x_1, x_2,....., x_n \mid \theta)$$

$$\sim \theta^{\sum_{i=1}^{n} x_i} \cdot (1-\theta)^{ns-\sum_{i=1}^{n} x_i}. \quad (13)$$

The right side of relations (11) is (up to a normalizing factor, independent of $\theta$) the density of the beta distribution

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b)} \theta^{a-1} \cdot (1-\theta)^{b-1} \quad (14)$$

with parameters $a = \sum_{i=1}^{n} x_i + 1$ and $b = ns - \sum_{i=1}^{n} x_i + 1$ respectively, where $\Gamma(z)$ - the known Euler gamma function (see. Example 1 a)).

Consequently, the family of conjugate prior distributions parameter $\theta$ (the probability of "success") observed the binomial distribution of the total population belongs to the class of beta - distribution (14).

It is known that the average value ($E\theta$) and variance ($D\theta$) beta distribution parameters and expressed in terms $a$ and $b$ of this distribution by the formulas:

$$E\theta = \frac{a}{a+b} = \theta_0, D\theta = \frac{ab}{(a+b)^2(a+b+1)} = \Delta^2.$$

Using these expressions we obtain as the solutions of the system of two equations relatively $a$ and $b$:

$$a = \frac{\theta_0^2(1-\theta_0)}{\Delta^2} - \theta_0, b = \left(\frac{\theta_0^2(1-\theta_0)}{\Delta^2} - \theta_0\right)\frac{1-\theta_0}{\theta_0}.$$

We define the posterior distribution, taking into account (10) and the expression of gamma density - distribution:

$$\tilde{p}(\theta \mid x_1, x_2,....., x_n) \sim \theta^{\alpha-1} e^{-\beta\theta} \cdot \theta^{n-1} \cdot e^{-\left(\sum_{i=1}^{n} x_i\right) \cdot \theta}.$$

From here, it is seen that the posterior distribution of parameter $\theta$ again obeys the gamma - distribution, but with the following parameters:

$$\tilde{\alpha} = \alpha + n, \tilde{\beta} = \beta + \sum_{i=1}^{n} x_i.$$

v) The random variable with unknown value of the parameter $\theta$ has a uniform distribution on the segment $[0;\theta]$. Check the conditions of existence of the conjugate priori distribution, to determine its accessories thereof Class of distributions and the specific values of the parameters of the distribution.

It is known that

$$f(x \mid \theta) = \begin{cases} \dfrac{1}{\theta} & for \quad 0 \le x \le 0, \\ 0 & for \quad x \notin [0;\theta]. \end{cases}$$

From equation (2) have:

$$L(x_1, x_2,....., x_n \mid \theta) = \left(\frac{1}{\theta}\right)^n, for \ x_{max}(n) = \max_{1 \le i \le n} x_i \le \theta. (15)$$

Therefore, within the overall view (6) we have: $m = 1$; $T(x_1, x_2,..., x_n) = x_{max}(n)$ - sufficient statistic that confirms the existence of a priori, conjugated with $L(x_1, x_2,....., x_n \mid \theta)$ the distribution of the parameter $\theta$. Now will update the belonging of this distribution to a certain class of distributions.

Taking into account (7), from (8) we obtain:

$$\hat{p}(\theta \mid x_1, x_2,....., x_n) \sim \hat{p}(\theta) \cdot L(x_1, x_2,....., x_n \mid \theta)$$

$$\sim \left(\frac{1}{\theta}\right)^{n-1} \quad for \ x_{max}(n) \le \theta. \quad (16)$$

The right side of the relations (16) is the density distribution of the Pareto of the form

$$p(\theta) = \begin{cases} \dfrac{\alpha \theta_{min}^{\alpha}}{\theta^{\alpha+1}} & for \quad \theta \ge \theta_{min} \\ 0 & for \quad \theta < \theta_{min} \end{cases} \quad (17)$$

with parameters $\alpha = n$ and some parameter shift

$$\theta_{min} \ge x_{max}(n).$$

Consequently, the family of conjugate prior distributions on the segment $[0;\theta]$ of parameter $\theta$ uniformly distributed random variable belongs to the Pareto distribution of the form (17).

We define a priori distribution, taking into account (15) and the expression density Pareto distribution (17) as a priori distribution

$$p(\theta; D) = \rho(\theta; \alpha; \theta_{min})$$

we have:

$$\tilde{p}(\theta \mid x_1, x_2,....., x_n) \sim \frac{\alpha \theta_{min}^{\alpha}}{\theta^{\alpha+1}} \cdot \frac{1}{\theta^n}$$

for $\max\{\theta_{min}; x_1, x_2,..., x_n\} \le \theta$.

It follows that the a priori distribution of the parameter $\theta$ is described, as well as a priori, by Pareto law (17), but with the following parameters:

$$\tilde{\alpha} = \alpha + n, \tilde{\theta}_{min} = \tilde{\theta}_{min} = \max\{\theta_{min}; x_1, x_2,..., x_n\}.$$

*Example 2*. A result of inspection 10 firms it was found that 6 firms have an average reliability, 3 firm - high reliability and 1 firm - low reliability. A priori reliability probabilities equal 0.5 , 0.3 , 0.2 respectively. Find a posteriori expectations of probability reliability of firms.

In this example, there are three events ( $m = 3$ ): firms have an average reliability (the first event - $\omega_1$ ), high reliability (the second event - $\omega_2$ ), low reliability (the third event - $\omega_3$ ). The probabilities of these events are random variables and mathematical expectations of probability reliability of firm ( $t_1, t_2, t_3$ ) for the Dirichlet a priori distribution equal: $t_1 = 0.5$ , $t_2 = 0.3$ , $t_3 = 0.2$ .

Considering that, $n_1 = 6$ , $n_2 = 3$ , $n_3 = 1$ , $N = 10$ and took $s = 1$ , the posterior probability of reliability expectations of firms are equal:

$$t_1^* = \frac{6 + 1 \cdot 0.5}{10 + 1} = 0.591,$$

$$t_2^* = \frac{3 + 1 \cdot 0.3}{10 + 1} = 0.3,$$

$$t_3^* = \frac{1 + 1 \cdot 0.2}{10 + 1} = 0.109.$$

Thus, after receiving the new statistical information probability of the reliability of firms are random, but with new mathematical expectations obtained above. In this information confirmed the a priori idea of the probability of firms of high reliability and changed the idea of the probabilities of low and medium of firms reliability. On one hand, if we take $s = 0$ , then the a posteriori information be fully determined only by the available additional information as analysis 10 firms. The data of conversion in this case do not depend on the choice of a priori probabilities. On the other hand, if the parameter *s* takes large values, the resulting statistics practically ceases to affect the posterior probabilities and the search for additional information loses its meaning.

Bayesian models have several advantages compared with the frequency models. One such advantage is that Bayesian models can in principle give some results, even if not at the sample data. This is due to the use of the a priori probability distribution, which in the absence of statistical data does not change, and calculated by Bayes' theorem the posteriori distribution coincides with the a priori.

## 4. Conclusions

In deciding Bayesian models allow to carry out additional experiments to clarify the states of nature. In practice, this refinement is mainly by collecting additional information, and by performing these experiments. Additional information about the profit and timely settlement with the budget allows us to refine the a priori probabilities of states of nature (for reliability).

At the same time, the design of experiments to determine whether it is expedient from the viewpoint of the cost of its implementation. To do this, you need to compare the expected additional income or benefits that be obtained by using additional information acquired as a result of the experiment, and the expected cost of the experiment. So, for example, with an analysis of the reliability the firm for information about the profit may demand an audit on which to spend certain means. The number of alternative courses of action in such problems increases as an opportunity to not only select one of the available alternatives, but also determine the feasibility of carrying out experiments or activities for additional information.

## References

[1] Tulupyev, A.L., Nikolenko S.I., Sirotkin A.V., *BAYESIAN NETWORKS: A Probabilistic Logic Approach (Russian)*, St. Petersburg: Nauka, 2006. 607 p.

[2] Morris, U.T, The science of management: A Bayesian approach.. M: Mir, 1971. 304 p.

[3] Friedrich Schreiber, Probability and Bayesian Statistics Plenum Press, The Extended Bayes-Postulate, Its Potential Effect on Statistical methods and Some Historical Aspects, New York, 1987, pp. 423-430.

[4] Jeffreys, Scientific Inference. 2nd ed. Cambridge University Press, 1957.

[5] De Groot M.H. Optimal Statistical Decisions: Wiley, 2004.