Mathematica Balkanica

Mathematical Society of South-Eastern Europe
A quarterly published by
the Bulgarian Academy of Sciences – National Committee for Mathematics

Mathematica
Balkanica

New Series Vol. 14, 2000, Fasc. 3-4

A Statistical Agent's Belief in Gaussian Model

Veska Noncheva

Presented by Bl. Sendov

Nowadays statistical software aims at creating intelligent statistical agents. These agents embody statistical expertise and knowledge that allows them to exhibit intelligent behaviour, to cooperate with users and other agents in problem solving and to reason under uncertainty.

The first aim of this article is to show an approach to modeling of the belief of a statistical agent in the normality of the population distribution. The agent's belief is a threshold function of the certainty with a multitude of values - the set of the states of the belief. The rule for defining the state of the agent's belief is based on: a function describing the utility of the state, a conditional distribution of the agent's certainty, the condition being the tests' results and a conditional distribution of the post-test result, the condition being the pre-test result, and also on threshold values of the pre-test and post-test results.

The second aim is to discuss an approach to optimizing decision rules determining the states of agent's belief. This approach takes into consideration a priori knowledge. It represents an application of Bayes' approach.

AMS Subj. Classification: 68T35, 68N05, 62C12

Key Words: computer science, artificial intelligence, agent's belief, statistical software, utility function

1. Introduction

The purpose of this work is to present an approach to modeling of the agent's belief in the particular case when the belief concerns the form of the probability distribution of the population under investigation.

"The reflection of reality can be true and untrue. In the process of communication people feel the necessity to express their attitude towards facts, phenomena, actions, etc. Sometimes they need to stress explicitly on their confidence (especially if there are doubts or vacillations in the interlocutor)." ([4])

We shall define the concept agent's belief in the following manner:

The agent's belief will be measured by a real number, called certainty. We say that the agent believes in the truth of a given statement if the certainty is not smaller than a number that is known in advance. That means that the agent's belief is a threshold function.

"The suspicion is an assumption connected with some doubts in the revealing of an unknown quantity. It manifests itself in the search for the truth." ([4])

"The certainty is connected with the authenticity, veracity, the possibility and the confidence that something has happened or will happen. It is both a state and a volitional category." ([4])

Formally, the suspicion and the certainty will be defined as possible values of the threshold function belief.

The rule for defining the state of the agent's belief is based on:

- a function describing the utility of the state,
- a conditional distribution of the agent's certainty, the condition being the tests' results and a conditional distribution of the post-test result, the condition being the pre-test result,
- threshold values of the pre-test and post-test results.

There are several techniques available for empirically assessing the mathematical form and the parameters of a utility function ([3], [6], [9]).

Both the conditional probability distribution of the agent's certainty on the values of the tests and the conditional probability distribution of the post-test result on the value of the pre-test must be stochastically increasing. This condition is met in many distributions widely used. For example, members of the one-parameter-exponential family (normal, binomial, Poisson, etc.) have a monotone likelihood ratio, which implies that the requisite posteriori distributions are stochastically increasing ([2]).

The task is set to have the expected utility maximized and thus the threshold values of the pre-test and post-test results are optimized.

2. Formulation of the task

The necessity for defining the agent's belief about the type of the probability distribution arises from the fact that the population distribution, which is at the root of many models of the mathematical statistics, is unknown. As a result of this insufficient knowledge comes the indefiniteness at the choice of the appropriate method for statistical analysis, i.e. the indefiniteness at the choice of the best behaviour of the agent, making the statistical analysis.

Let T be a random variable which has a probability distribution on the set $\Omega_t = [0,1]$. Let $t \in \Omega_t$ be a possible value of T. We call the random variable T *agent's certainty.*

Let Ω_t be divisible by $\{A_k = [t_k, t_{k+1}], t_k \in \Omega_t, t_{k+1} \in \Omega_t, k = 0, 1, ..., n-1\}$, i.e. $A_i \cap A_j = \emptyset$ when $i \neq j$, and $\sum_{k=0}^{n-1} A_k = \Omega_t$.

Let δ be a categorical variable which can take on values from the set $D = \{a_0, a_1, ..., a_{n-1}\}$.

We are going to define the random variable *Bel* as a function of the random variable T in the following manner:

$$Bel(T) = \begin{cases} a_0 & \text{if} \quad t_0 \leq T < t_1, \\ a_1 & \text{if} \quad t_1 \leq T < t_2, \\ \dots & \\ a_{n-1} & \text{if} \quad t_{n-1} \leq T < t_n. \end{cases}$$

The random variable *Bel(T)* will be called *agent's belief.* The values $a_0, a_1, ..., a_{n-1}$, which the random variable *Bel* can take on, will be called *decisions*, or *states of the belief.* We shall call the set D *a set of the possible states of the belief.*

Consider the agent's belief about the normality of the population distribution. In this case $D = \{a_0, a_2\}$, where:

a_0 means that the agent rejects the assumption about the normality of the distribution, which describes the population under consideration.

a_2 means that the agent is certain that the distribution, which describes the population under consideration, is normal and it will use this assumption in the statistical analysis.

The agent's belief about the normality of the population distribution has the following form:

$$Bel(T) = \begin{cases} a_0 & \text{if} \quad 0 \leq T < t_c, \\ a_2 & \text{if} \quad t_c \leq T < 1, \end{cases}$$

where t_c is a known constant. For example, $t_c = 0,95$.

Our objective is to determine the value of the function $\delta = Bel(T)$. In this case we say that the agent makes a decision $\delta \in D$. However, the agent's certainty is an unobservable random variable.

The agent's belief in Gaussian probability model (i.e. in the normality of the probability distribution) must be based on the knowledge of the mechanism of the phenomenon under investigation. But if the phenomenon under investigation is unknown, the agent can make its own choice about the probability distribution after it has tested a Goodness-of-fit hypothesis. We shall elucidate this statement.

Let $X_1, X_2, ..., X_n$ be a random sample from the distribution of the random variable, describing the population. We would like to know whether it is rational to examine this random variable as normally distributed.

Let $x_1, x_2, ..., x_n$ be the experimental values of $X_1, X_2, ..., X_n$, respectively.

We propose the statistical hypothesis H_0 that the experimental data are values of a random variable with normal $N(\mu, \sigma^2)$ distribution where μ and σ^2 are determined.

A statistical test is the rule by means of which we make a decision to reject the statistical hypothesis H_0 or not to reject it.

A significance level a of the statistical test is the maximum probability for the rejection of the null hypothesis H_0, when it is true.

Let $U = u(X_1, X_2, ..., X_n)$ be the statistic of the optimal statistical test, the probability distribution of which is known on condition that the null hypothesis is true. Let $u = u(x_1, x_2, ..., x_n)$ be the value of the statistic. Then $p(x_1, x_2, ..., x_n) = P\{U \geq u(x_1, x_2, ..., x_n)/H_0\}$ is called p-value of the test statistic. A result is statistically significant when the p-value is less than the preset value of α.

Then the statistical test has the following form:

If $p(x_1, x_2, ..., x_n) < \alpha$, where $\alpha = 1 - t_c$ is the desired significance level of the statistical test, then the null statistical hypothesis is rejected in favour of the alternative hypothesis. Otherwise, we have no reason to reject the null hypothesis, i.e. we think that the data we have do not give reasons to reject the hypothesis for normality.

After the Goodness-of-fit hypothesis has been tested at a significance level $\alpha = 1 - t_c$, the agent can make its preliminary choice about the probability distribution.

However, if the sample size is not big enough or the results from the measurements are located in a sufficiently small interval of change of the random variable that describes the population, then the experimental data can concur well with different probability models. That is why one must not attach much importance to the positive result of the testing of hypotheses.

We emphasize once again that the choice of the appropriate probability model must be based above all on the understanding of the mechanism of the phenomenon under examination.

When the agent makes a decision about the form of the probability distribution, he can ask the user and make use of its expertise and intuition.

If the user does not know the mechanism of the phenomenon under investigation either, then the user can make his choice of distribution taking the following considerations into account:

- The histogram.

  We must bear in mind that a histogram can be the resemblance of several distributions. For example, by the histogram it is practically impossible to differentiate between the lognormal distribution and Weibull's distribution, even at a large sample size.

- The graphic representation of the empirical cumulative function on probability paper.

  But there is no quantity criterion for the possible deviation of the values of the empirical cumulative function from the straight line.

- The sample coefficients of asymmetry and excess.

  If for a sample there have been calculated the point estimates of the asymmetry and of the excess, as well as their mean quadratic deviations, then we assume that the empirical cumulative function concurs with the theoretical one if the sample coefficients of asymmetry and excess differ in their absolute value from their mathematical expectations no more than the mean quadratic deviation trebled.

  When making a decision in this manner, the form of the curve is taken into account. Yet the rule of the three sigmas is an empirical rule.

- The results from the Goodness-of-fit test.

  A formal statistical test can be used to test whether the distribution of the data differs significantly from a Gaussian distribution. However, with few data points, it is difficult to tell whether the data are Gaussian by inspection, and the formal test has little power to discriminate between Gaussian and non-Gaussian distributions.

Besides the main Goodness of fit tests – χ^2 test and Kolmogorov's test, there are also others, e.g. Missies-Smirnoff's test which, unlike the χ^2 test, does not require combining the numerical values in intervals; in other words it makes a better use of the information that is contained in the sample. There is also a test for normality, using a sufficiently big number of small volume samples.

Thus, it is not always easy to decide whether a sample comes from a Gaussian population. The user remembers that what matters is the distribution of the overall population, not the distribution of his sample. In deciding whether a population is normal, he looks at all available data, not just the data in the current experiment. However, when the scatter comes from the sum of numerous sources (with no one source contribution accounting for most of the scatter), the user expects to find a roughly Gaussian distribution.

When making a decision about the form of the probability distribution following one or more of the above-discussed ways, the skilled statistician can conclude, "I am 99% certain that the distribution of the population under investigation is normal." Another user might insist that he is only 80% certain in the normality of the population distribution. Yet another one realizes that he cannot define the form of the distribution and he claims that by the same probability it could be normal or it could as well not be normal. The degree of certainty of these three users can be represented respectively by the numbers 0.99, 0.8, 0.5.

Consequently, the agent will make use of a pre-test and a post-test when making a decision about the normality of the symmetrical continuous population.

Usually, as a pre-test there is used a statistical test about the form of the probability distribution, and after that the user is asked about his opinion. The results of those two tests are respectively $1-p$, where p is the p-value of test statistic and the user's degree of certainty, represented as numbers in the interval [0,1].

Sometimes, e.g. in regression analysis, the user is first asked if the population distribution is normal, then the model is built and afterwards the residuals from the estimated regression model are tested for normality.

It is natural to expect that the high values of the pre-test results will lead to lower requirements of the post-test results.

3. Model

Let us mark the pre-test results with X and the post-test results with Y. Let us assume that X and Y are continuous random variables with probability distribution respectively on $\Omega_x = [0,1]$ and $\Omega_y = [0,1]$.

By the results of the observations x and y, we must make a decision $\delta = \delta(x,y) \in D$. The function $\delta(x,y)$ defined on the set of possible results from the observations and accepting values in the space D of the possible solutions is called *a decision rule*, or *a rule for making a decision.*

Let T be the agent's certainty, i.e. a continuous random variable with a probability distribution on the set $\Omega_t = [0,1]$, and it cannot be observed.

Also suppose that the relation between the random variables X, Y and T can be represented by the joint density function $f(x,y,t)$. The rule for making a decision $\delta(x,y)$ determines for each possible realization (x,y) of the random vector (X,Y) which state $a_j, j = 0,1,2$ will be appropriated for the agent's belief.

Let t be the true value of the agent's certainty and a decision $d \in D$ has been made.

Let us designate with $u(d,t)$ the utility of making a decision d.

If we use the decision rule $\delta(x,y)$then the expected utility is $U(\delta/t) = Eu(\delta(x,y),t)$.

The function $U(\delta/t)$ is called expected utility at the use of the decision rule $\delta(x,y)$ on condition that the true value of the agent's certainty is t.

It seems reasonable to choose this rule for making a decision which the maximum expected utility corresponds to. We are looking for a decision rule, which is the best in this sense.

For each decision rule δ we designate $U(\delta) = \int\limits_{\Omega_t} U(\delta/t)\mu(t)dt$.

The decision rule δ^* is called Bayes' if $U(\delta^*) = \max U(\delta)$.

Let us go back to our task of defining the agent's belief in the normality of the population under investigation. We shall add a new element a1 to the space of the decisions, where the state means that the agent suspects that the distribution of the population is normal.

Then $D = \{a_0, a_1, a_2\}$, where:

a_0 means that the agent rejects the assumption of the normality of the distribution which describes the population under investigation;

a_1 means that the agent suspects that the distribution is normal. In the process of the statistical analysis it will make use only of tests which are not sensitive to moderate deviations from the assumption for normality. An example of such a robust test is the t-test;

a_2 means that the agent is convinced that the distribution is normal. It will also use statistical tests, which are sensitive even to moderate deviations from the assumption for normality. Such tests are, for example, Pearson's, Fisher's and Bartlett's tests for equality of variances.

We shall divide the users' population into $g = 2$ subpopulations:

- a subpopulation of the users with statistical expertise which will be called a subpopulation of the skilled statisticians, and
- a subpopulation of the statisticians-beginners that will be called a subpopulation of the naive statisticians.

We shall also assume that the relation between the measured results from the pre-test X, the measured results from the post-test Y and the certainty T in the subpopulation i can be described by the joint density function.

It is intuitively clear that the high values of the pre-test results lead to lower requirements of the post-test results. That is, the received preliminary information influences the decision rules. Decision rules in which the post-test decisions are functions of the pre-test results will be called weak rules. The weak rule for making a decision d can be defined as follows:

$$\begin{aligned}\{(x,y):\delta(x,y)=a_0\}&=A_i\times[0,1]\\ \{(x,y):\delta(x,y)=a_1\}&=A_i^c\times B_i(x)\\ \{(x,y):\delta(x,y)=a_2\}&=A_i^c\times B_i^c(x),\end{aligned} \tag{1}$$

where A_i and A_i^c are sets of values of the pre-test leading to the rejection and to the acceptance of the statement for normality respectively. $B_i(x)$ and $B_i^c(x)$ are sets of values of the post-test leading respectively to the rejection and to the acceptance of the statement for normality.

It is natural to consider rules for making a decision, which have a monotonous form. For example, the agent is convinced that the distribution of the population is normal if the value of the variable, representing the certainty, is not smaller than a fixed in advance threshold value (probability) otherwise the agent rejects the assumption for normality.

Consequently, the problem for making a decision about the type of the probability distribution of the population under consideration is reduced to finding two threshold points (probabilities) x_c and y_c for X and Y respectively which are optimum with respect to Bayes' approach, i.e. $A_i = [0, x_c)$, $B_i = [0, y_c)$.

The existence of users with different statistical competence leads to various values of the threshold points x_c and y_c which we shall designate with x_{ci} and y_{ci}, respectively. The threshold point t_c on T is one and the same for each population of users. It is fixed in advance and known to the agent making a decision. Usually t_c=0,95.

Then the weak monotonous rules for making a decision will be defined as follows:

$$\delta(x,y)=\begin{cases} a_0 & \text{if } X<x_{ci}\\ a_1 & \text{if } X\geq x_{ci}, Y<y_{ci}(x)\\ a_2 & \text{if } X\geq x_{ci}, Y\geq y_{ci}(x),\end{cases} \tag{2}$$

where $a_j, j = 0,1,2$ are the states of the agent's belief, x_ci and y_ci are the threshold points for X and Y in the subpopulation i.

4. Utility structures

A utility function is called the function $u_{ji}(t)$ which describes the utility of the result from the appropriation of the state $a_j, (j = 0,1,2)$ of the agent's belief for a case with a user from subpopulation i, in which case the real value of the agent's certainty is t.

We shall determine two types of utility functions. Our aim is to define realistic utility functions. We would like the utility to be changed smoothly from correct to incorrect decision without jumps in the points close to the threshold value of agent's certainty t_c. For this purpose we are going to use a continuous one. A model for making a decision about the type of the population distribution, making use of a linear utility function is described in [7]. It is natural for us to want the utility function to be limited also in the remote from the t_c points, the utility to increase evenly and evenly to decrease to the respective boundaries.

Consider a utility function which is a continuous limited function of the true value of the agent's certainty. The Probit models are suitable for this purpose.

In accordance with Berhold's ideas [1], we define a utility function that has the following form:

$$u_{ij}(t) = \begin{cases} \phi[(\mu_{0i} - t)/\sigma_{0i}] & \text{if } j = 0 \\ \phi[(t - \mu_{1i})/\sigma_{1i}] & \text{if } j = 1 \\ \phi[(t - \mu_{2i})/\sigma_{2i}] & \text{if } j = 2, \end{cases} \tag{3}$$

where $\phi[.]$ is the standard Gaussian distribution function, μ_{ji} and $\sigma_{ji} > 0, i = 1, 2, j = 0, 1, 2$, are the mean and the standard deviation of the Gaussian distribution, respectively.

This structure will be called the Gaussian utility structure. The Gaussian distribution functions are symmetrical with respect to 0.5 in the meaning that $\phi(\mu + x) - 0.5 = 0.5 - \phi(\mu - x)$.

If we want our utility function to increase smoothly from its lower boundary and to aim quickly at its upper boundary, we shall use the complementary log-log function $\pi(x) = 1 - \exp(-\exp(\alpha + \beta x))$. It is an asymmetrical function and when $\beta > 0$ it aims faster at its upper boundary than at its lower one.

Following this approach we shall use the utility structure with the following form:

$$u_{ji}(t) = \begin{cases} 1 - \exp(-\exp(b_{0i}(t_c - t)) & \text{if } j = 0, \\ 1 - \exp(-\exp(b_{1i}(t - t_c)) & \text{if } j = 1, \\ 1 - \exp(-\exp(b_{2i}(t - t_c)) & \text{if } j = 2, \end{cases} \tag{4}$$

where $b_{ji} > 0, j = 0, 1, 2$, and $t_c \in [0, 1]$ is a fixed number.

The utility function, defined in (4) will be called *complementary log-log utility structure.*

The values of $b_{ji}, i = 1, 2, ..., g, j = 0, 1, 2$ can be different for the different subpopulations of users. The conditions $b_{0i} > 0$ and $b_{2i} > 0$ are equivalent to the statement that the utility is strictly decreasing in the cases when the agent rejects the assumption for normality, and it is strictly increasing in the cases when the agent is convinced that the distribution which describes the population under investigation is normal.

Since for narrower assumptions about the form of the population distribution are used stronger methods, i.e. stronger results are received from the statistical analysis, we shall assume that $u_{1i}(t)$ is an increasing function from t, i. e. is $b_{1i} > 0$.

Let us suppose that h is the population of the skilled statisticians and i is the population of the naive statisticians. It is natural for a naive statistician to rely more on the opinion and advice of the agent while making the statistical analysis than a skilled statistician does. Then an incorrect decision will have worse consequences for the population i than for h. Or, which is almost the same, the correct decision is more valuable for the population i than for the population h. That is why we shall choose $b_{ji} > b_{jh}, j = 0, 2$.

5. Sufficient conditions for the monotony of the weak rules

We shall establish the optimum rules for making a decision by maximizing the expected utility. The limitation to monotonous rules is correct if there are not non-monotonous rules with bigger expected utility.

In this paragraph the particular form of the utility function is not used. It is sufficient to assume that it is continuous and bounded.

Let us consider the subpopulation i of the users. We find for rules (1) the expected utility for a fixed value of i.

$$Eu_i(A_i^c, B_i^c(x)) = \int_{A_i} \int_0^1 u_{0i}(t).w_i(x,t)dtdx$$

$$+ \int_{B_i(x)} \int_0^1 u_{1i}(t).f_i(x,y,t)dtdydx + \int_{A_i^c} \int_{B_i^c(x)} \int_0^1 u_{2i}(t).f_i(x,y,t)dtdydx,$$

where $w_i(x,t)$ is the joint probability density function of X and T in the i subpopulation.

We mark with $E_i(g(T)/x)$ the regression function of $g(t)$, where it is fixed that $X = x$. We mark with $k_i(x,y)$ the joint probability density function of X and Y in the i subpopulation. We mark with $E_i(g(T)/x,y)$ the regression function of $g(T)$, where $X = x$ and $Y = y$ are fixed. Then,

$$Eu_i(A_i^c, B_i^c(x)) = E_i\{u_0(T)\} + \int_{A_i^c} E_i\{[u_1(T) - u_0(T)]/x\}.q_i(x)dx$$

$$+ \int_{A_i^c} \int_{B_i^c(x)} E_i\{[u_2(T) - u_1(T)]/x, y\}.k_i(x, y)dydx.$$

We shall represent $k_i(x, y) = h_i(y/x).q_i(x)$. Then

$$Eu_i(A_i^c, B_i^c(x)) = E_i\{u_0(T)\} + \int_{A_i^c} \{E_i\{[u_1(T) - u_0(T)]/x\}$$

$$+ \int_{B_i^c(x)} E_i\{[u_2(T) - u_1(T)]/x, y\}h_i(y/x)dy\}q_i(x)dx. \tag{5}$$

Therefore the problem is reduced to the maximizing of (5) for each value of $i, i = 1, 2, ..., g$.

We find an upper boundary of the expected utility. For this purpose we make use of almost the obvious statement *Suppose that* $\int |f(x)|\, dx < \infty$ *is fulfilled for the function* $f(x)$. *Then for each set of values of* x *the inequality* $\int_S f(x)dx \leq \int_{S_0} f(x)dx$holds, where $S_0 = \{x : f(x) \geq 0\}$.

From (5) it follows that for each $B^c(x)$ when A^c is fixed, the following inequality holds:

$$Eu(A^c, B^c(x)) \leq E\{u_0(T)\} + \int_{A^c} \{E\{[u_1(T) - u_0(T)]/x\}$$

$$+ \int_{B_0^c(x)} E\{[u_2(T) - u_1(T)]/x, y\}.h(y/x)dy\}.q(x)dx, \tag{6}$$

where

$$B_0^c(x) = \{y : E\{[u_2(T) - u_1(T)]/x, y\} \geq 0\}. \tag{7}$$

After applying the statement once again, from (6) we get that for each A^c the following inequality holds:

$$Eu(A^c, B^c(x)) \leq E\{u_0(T)\} + \int_{A_0^c} \{E\{[u_1(T) - u_0(T)]/x\}$$

$$+ \int_{B_0^c(x)} E\{[u_2(T) - u_1(T)]/x, y\}.h(y/x)dy\}.q(x)dx,$$

where

$$A_0^c = \{x : E\{[u_1(T) - u_0(T)]/x\} + \int_{B_0^c(x)} E\{[u_2(T) - u_1(T)]/x, y\}.h(y/x)dy \geq 0\},$$

(8)

If the function

$$E\{[u_2(T) - u_1(T)]/x, y\} \tag{9}$$

from (7) is increasing on y for each x and the function

$$E\{[u_1(T) - u_0(T)]/x\} + \int_{B_0^c(x)} E\{[u_2(T) - u_1(T)]/x, y\}h(y/x)dy \tag{10}$$

from (8) is increasing on x for each y, then the expected utility (see (5)) is maximized on the sets $A_0^c = [x_c, 1]$ and $B_0^c(x) = [y_c(x), 1]$, where x_c and $y_c(x)$ are the roots respectively of the equations from (8) and (7).

A weaker condition for maximizing (5) on the sets $A_0^c = [x_c, 1]$ and $B_0^c(x) = [y_c(x), 1]$ is the functions (9) and (10) to change their sign from negative to positive only once.

Definition. We say that the density function $f(t/x)$ is *stochastically increasing in* x, if from $x_1 \geq x_2$ follows that the distribution function satisfies $F(t/x_1) \leq F(t/x_2)$, where x_1 and x_2 are realizations of the random variable X.

Lemma 1. *Let T and Z be random variables with realizations t and z, respectively. For each increasing function $k(t)$ the conditional mathematical expectation $E[k(T)/z]$ is an increasing function in z if and only if the conditional density $f(t/Z = z)$ is stochastically increasing in z.*

The sufficient condition is proved in [5]. The necessary condition is proved in [2].

Lemma 2. *If the conditional density $v(t/x, y)$ of T on condition that $X = x$ and $Y = y$ is stochastically increasing in x and y, and the conditional density $h(y/x)$ of Y on condition that $X = x$ is stochastically increasing in x, then the marginal conditional density $p(t/x) = \int_0^1 v(t/x, y).h(y/x)dy$ is stochastically increasing in x.*

Lemma 3. *If the function $k(x, y)$ is strictly increasing on x and y, then the dependence of y on x defined by means of $(x, y) : k(x, y) = c, c \in R$ is a decreasing function on x.*

Lemma 2 and Lemma 3 are proved in [10].

Theorem. *Let the following conditions hold:*

(11) *The functions $u_j(t) - u_{j-1}(t), j = 1, 2$ are strictly increasing in t.*

(12) *The conditional distribution function $V(t/x, y)$ is strictly decreasing in x and y for each t.*

(13) *The conditional distribution function $H(y/x)$ is strictly decreasing in x for each y.*

Then the optimum weak rule for making a decision for normality of the probability distribution of the population being investigated is monotonous.

For the special case when the utility function has a linear form, the theorem is proved in [10]. We are going to prove it by analogy for a continuous and bounded utility function.

Proof. From conditions (11) and (12), and Lemma 1, it follows that $E\{[u_2(T) - u_1(T)]/x, y\}$ is increasing in y for each x and it is increasing in x for each y. Therefore, the set $B_0^c(x)$ takes the form $[y_c(x), 1]$ for all values of x. From theorem's conditions, Lemma 1 and Lemma 2, it follows that $E\{[u_1(T) - u_0(T)]/x\}$ is an increasing function in the point x. $E\{[u_2(T) - u_1(T)]/x, y\}$ is increasing on x and y, and is non-negative when $y \geq y_c(x)$ for each x (see (7)), and $h(y/x) \geq 0$.

Let $x_2 > x_1$. Then from Lemma 3 it follows that

$$\int_{y_c(x_2)}^{1} E\{[u_2(T) - u_1(T)]/x_2, y\}h(y/x_2)dy$$

$$- \int_{y_c(x_1)}^{1} E\{[u_2(T) - u_1(T)]/x_1, y\}h(y/x_1)dy$$

$$> \int_{y_c(x_1)}^{1} E\{[u_2(T) - u_1(T)]/x_2, y\}h(y/x_2)dy$$

$$- \int_{y_c(x_1)}^{1} E\{[u_2(T) - u_1(T)]/x_1, y\}h(y/x_1)dy$$

$$> \int_{y_c(x_1)}^{1} E\{[u_2(T) - u_1(T)]/x_1, y\}h(y/x_2)dy$$

$$- \int_{y_c(x_1)}^{1} E\{[u_2(T) - u_1(T)]/x_1, y\}.h(y/x_1)dy$$

$$= \int_{0}^{1} I_{[y_c(x_1),1]}(y).E\{[u_2(T) - u_1(T)]/x_1, y\}.[h(y/x_2) - h(y/x_1)]dy > 0,$$

where $I_{[y_c(x),1]}$ is the indicator of the set $[y_c(x), 1]$. We used the fact that the density $h(y/x)$ is stochastically increasing in x (see (13)).

Hence

$$E\{[u_1(T) - u_0(T)]/x\} + \int_{y_c(x)}^{1} E\{[u_2(T) - u_1(T)]/x, y\}h(y/x)dy$$

is an increasing function. Therefore, the set A_0^c takes the form $[x_c, 1]$.

The theorem is proved. ∎

Consider the Gaussian utility structure. Then

$$u_1(t) - u_0(t) = \varphi(\frac{t - \mu_1}{\sigma_1}) - \varphi(\frac{\mu_0 - t}{\sigma_0})$$

and

$$\frac{d}{dt}[u_1(t) - u_0(t)] = \frac{1}{\sigma_1}\varphi(\frac{t - \mu_1}{\sigma_1}) + \frac{1}{\sigma_0}\varphi(\frac{\mu_0 - t}{\sigma_0}) \geq 0$$

$\forall t \in [0, 1]$, where we have designated the density function of the standard normal distribution with $\varphi(.)$. Then,

$$u_2(t) - u_1(t) = \varphi(\frac{t - \mu_2}{\sigma_2}) - \varphi(\frac{t - \mu_1}{\sigma_1})$$

and

$$\frac{d}{dt}[u_2(t) - u_1(t)] = \frac{1}{\sigma_2}\varphi(\frac{t - \mu_2}{\sigma_2}) - \frac{1}{\sigma_1}\varphi(\frac{t - \mu_1}{\sigma_1}).$$

The roots of the equality

$$\frac{1}{\sigma_2^2}e^{-\frac{1}{2}\frac{1}{\sigma_2^2}(t-\mu_2)^2}-\frac{1}{\sigma_1^2}e^{-\frac{1}{2}\frac{1}{\sigma_1^2}(t-\mu_1)^2}=0$$

are

$$t_1=\frac{\sigma_1^{-2}\mu_1-\sigma_2^{-2}\mu_2-\sqrt{(\frac{\mu_1-\mu_2}{2})^2+(\sigma_2^2-\sigma_1^2)(\ln\sigma_2-\ln\sigma_1)}}{\sigma_1^{-2}-\sigma_2^{-2}}$$

and

$$t_2=\frac{\sigma_1^{-2}\mu_1-\sigma_2^{-2}\mu_2+\sqrt{(\frac{\mu_1-\mu_2}{2})^2+(\sigma_2^2-\sigma_1^2)(\ln\sigma_2-\ln\sigma_1)}}{\sigma_1^{-2}-\sigma_2^{-2}}.$$

Then the following four corollaries hold.

Corollary 1. *For the Gaussian utility structure (3), where $\sigma_{2i}>\sigma_{1i}>0, i=1,2$, the optimum weak rule for making a decision for the normality of the probability distribution is monotonous if:*

- $1<t_1$ *or* $t_2<0$, *where*

$$t_1=\frac{\sigma_1^{-2}\mu_1-\sigma_2^{-2}\mu_2-\sqrt{(\frac{\mu_1-\mu_2}{2})^2+(\sigma_2^2-\sigma_1^2)(\ln\sigma_2-\ln\sigma_1)}}{\sigma_1^{-2}-\sigma_2^{-2}}$$

  and

$$t_2=\frac{\sigma_1^{-2}\mu_1-\sigma_2^{-2}\mu_2+\sqrt{(\frac{\mu_1-\mu_2}{2})^2+(\sigma_2^2-\sigma_1^2)(\ln\sigma_2-\ln\sigma_1)}}{\sigma_1^{-2}-\sigma_2^{-2}},$$

- *the conditional distribution function $V(t/x,y)$ is strictly decreasing in x and y for each t.*
- *the conditional distribution function $H(y/x)$ is strictly decreasing in x for each y.*

Corollary 2. *For the Gaussian utility structure (3), where $\sigma_{1i}>\sigma_{2i}>0, i=1,2$, the optimum weak rule for making a decision for the normality of the probability distribution is monotonous if:*

- $t_1<0<1<t_2$, *where*

$$t_1=\frac{\sigma_1^{-2}\mu_1-\sigma_2^{-2}\mu_2-\sqrt{(\frac{\mu_1-\mu_2}{2})^2+(\sigma_2^2-\sigma_1^2)(\ln\sigma_2-\ln\sigma_1)}}{\sigma_1^{-2}-\sigma_2^{-2}},$$

$$t_2 = \frac{\sigma_1^{-2}\mu_1 - \sigma_2^{-2}\mu_2 + \sqrt{(\frac{\mu_1-\mu_2}{2})^2 + (\sigma_2^2 - \sigma_1^2)(\ln\sigma_2 - \ln\sigma_1)}}{\sigma_1^{-2} - \sigma_2^{-2}},$$

- *the conditional distribution function $V(t/x, y)$ is strictly decreasing in x and y for each t.*
- *the conditional distribution function $H(y/x)$ is strictly decreasing in x for each y.*

Corollary 3. *For the Gaussian utility structure (3), where $\sigma_{1i} = \sigma_{2i}, i = 1, 2$, the optimum weak rule for making a decision for the normality of the probability distribution is monotonous if:*

- $\mu_{1i} < \mu_{2i}$ *and* $\frac{\mu_{1i}+\mu_{2i}}{2} < 0$, *or* $\mu_{1i} > \mu_{2i}$ *and* $\frac{\mu_{1i}+\mu_{2i}}{2} > 1$, *where* $i = 1, 2$,
- *the conditional distribution function $V(t/x, y)$ is strictly decreasing in x and y for each t,*
- *the conditional distribution function $H(y/x)$ is strictly decreasing in x for each y.*

Corollary 4. *For the complementary log-log utility structure (4) the optimum weak rule for making a decision for the normality of the probability distribution is monotonous if:*

- $b_{2i} > b_{1i}$,
- *the conditional distribution function $V(t/x, y)$ is strictly decreasing in x and y for each t.*
- *the conditional distribution function $H(y/x)$ is strictly decreasing in x for each y.*

It is natural for the utility function to take on bigger values when the agent is certain that the distribution of the population under investigation is normal, unlike the cases when the agent suspects that the distribution is normal. Hence, the condition $b_{2i} > b_{1i}$ is a natural assumption.

The proof of Corollary 4 follows directly from the fact that when $b_{2i} > b_{1i}$ the function $u_2(t) - u_1(t)$ changes its sign from negative to positive exactly once in the interval $[0, 1]$ as well as from the properties of the mathematical expectation.

Let us now, summarize the result obtained:

The sets $B_0^c(x)$ and A_0^c are defined as follows: First, for all values of x, the sets $B_0^c(x)$ are defined by means of (7). Then the set A_0^c is defined by means of (8). Hence the optimum weak rules must be calculated in this order.

6. Conclusions

In this paper an approach to modeling of the belief of a statistical agent in the normality of the probability distribution of the population under investigation and an approach to optimizing decision rules determining the states of agent's belief are presented.

Further work on agent's belief will be published in [8]. The agent's belief in the probability model is presented via three items $(\mathcal{B}, \mathcal{D}, \mathcal{U})$. $\mathcal{B}$ is a Bayesian network, presenting the probability structure of the agent's belief problem, $\mathcal{D}$ is a decision network of the agent's belief state, and $\mathcal{U}$ is a utility network, presenting the utility structure of the agent's belief problem. The decision for the agent's belief state can be made via propagation in $\mathcal{D}$. The agent's belief state can be optimized via propagation in $\mathcal{U}$, using at the same time $\mathcal{B}$.

These approaches will be applied in a Multi-agent system for statistical hypotheses testing.

References

[1] M. H. B e r h o l d. The use of distribution functions to represent utility functions. *Management Science*, **19**, No 7, 1973, 825-829.

[2] D. T. C h u a n g, J. J. C h e n, M. R. N o v i c k. Theory and practice for the use of cut-scores for personnel decisions. *Journal of Educational Statistics*, **6**, 1981, 129-152.

[3] C. E. D a v i s, J. H i c k m a n, M. R. N o v i c k. *A primer on decision analysis for individually prescribed instruction.* Iowa City: Publication of the Research and Development Division, The American College Testing Program, 1973.

[4] B. I. K r a s t e v. *Communication Grammer* (in Bulgarian). Sofia, 1990.

[5] E. L. L e h m a n n. *Testing Statistical Hypotheses.* Wiley, N. York, 1959.

[6] R. D. L u c e, H. R a i f f a. *Games and Decisions.* John Wiley and Sons, N. York, 1957.

[7] V. N o n c h e v a. An approach to optimizing the agent's belief about the probability distribution of a population. *Scientific Works of Plovdiv University "Paissii Hilendarski"*, **33**, No 3, 1998.

[8] V. N o n c h e v a. Agent belief: presentation, propagation, and optimization. *Informatica*, To appear.

[9] B. J. V r i j h o f, G. J. M e l l e n b e r g h, W. P. v a n d e n B r i n k. Assessing and studying utility functions in psychometric decision theory. *Applied Psychological Measurement*, **7**, No 3, 1983, 341-357.

[10] H. J. V o s. *Simultaneous Optimization of Test-based Decisions in Education.* PhD. Thesis, University of Twente, Netherlands, 1994.

Faculty of Mathematics and Informatics
University of Plovdiv
24 Tsar Assen St., 4000 Plovdiv, BULGARIA
e-mails: nonchev@plovdiv.techno-link.com, wesnon@ulcc.uni-plovdiv.bg

Received: 09.12.1999