Considering information uncertainty when assessing risk in Bayesian Belief Network

Yu E Gagarin, U V Nikitenko and M A Stepovich

Bauman Moscow State Technical University (National Research University), Kaluga Branch, 2 Bazhenov Street, Kaluga, 248000 Russia
Tsiolkovsky Kaluga State University, 26 Stepan Razin Street, Kaluga, 248023 Russia

E-mail: g_ug@mail.ru

Abstract. The problem of estimating the average risk in Bayesian Belief Networks is considered, taking into account the uncertainty of the initial data. For estimation of parameters of functional dependences taking into account errors in values of functions and arguments it is offered to use methods of the confluent analysis. The necessity of obtaining interval estimates of functional dependencies and average risk to increase the reliability of decision-making is shown. The Bayesian Belief Network is modeled taking into account random errors.

1. Introduction

For mathematical modeling and engineering design, obtaining information about model parameters from observations of features is one of the main goals. The set of possible parameters in the simulation model corresponds to some set of observed or desired values. Bayesian networks can serve as a means of describing the cause-effect relationships of model parameters based on incoming informative signals. In Bayesian Belief Network, random events are connected by causal relationships, and such networks are used in situations that are characterized by inherited uncertainty. Description of situations by the method of cause and effect allows estimating the probability of events. Bayesian confidence networks are structures usually represented graphically as a directed acyclic graph and conditional probability tables for graph nodes corresponding to certain variables. The process of working with them is to perform two basic operations: training (formation of conditional probability tables) Bayesian Belief Networks on the basis of available data on network variables and direct use to calculate the various probabilities associated with network variables. In many cases, events can take any state from some range. That is, a network node is a continuous random variable whose possible state space is the entire range of its admissible values [1]. A random variable can have several States and is determined by a priori probability. The article deals with the situation when the vertices of a graph are connected by conditional probability densities, the parametric form of which is known. In this case, the initial data are samples of feature values from these distributions.

2. Estimation of parameters taking into account errors in the values of functions and arguments

Initial data are usually obtained as a result of experiments. And like any measurements, the original data contains random errors that need to be taken into account. Sources of measurement errors can be instrumental inaccuracy in the measurement of physical quantities, incomplete observations and preference of one data set over another. Known methods of estimating the parameters of conditional
probability densities for a given sample do not allow fully ing take into account random errors of the initial data. For example, the least squares method does not take into account the error of the conditional probability densities arguments. It is assumed that the values of the features are deterministic values.

Let there be a set of events $A_j$, $j = 1, M$. The parametric form of conditional probability densities

$$P(x \mid A_j)$$

for each event is known and the a priori probabilities of events are known $P(A_j)$.

Denote $\eta_j(x, \Theta) = P(x \mid A_j)P(A_j)$, where $\Theta$ – vector of unknown parameters.

The raw data are the results of specific experiments, and like any measurements contain random errors that need to be taken into account. In reality, the values of the signs $x_i$, $i = 1, n$, are not true and include random errors:

$$\begin{cases}
y_i = \eta(\xi_i, \Theta) + \epsilon_i, \\
x_i = \xi_i + \delta_i,
\end{cases} \quad i = 1, n,$$

(1)

where $\epsilon_i$ и $\delta_i$ – the error function value and argument, respectively; $\xi_i$ - unknown (true) values of the characteristics.

Assume that measurement errors $\epsilon_i$ и $\delta_i$ – normally distributed random variables with zero mean values, with variances $\sigma^2(y_i)$ and $\sigma^2(x_i)$ respectively, and the correlation coefficient $\rho_j = 0$.

To account for errors in the values of functions and arguments when evaluating parameters, there are a number of methods [2-4], which are based on different approaches. To obtain unbiased estimates of the model parameters (1), we use the methods of confluent analysis [5-6].

Parameter estimates $\hat{\Theta}$ are derived from the minimum functional condition:

$$F = \frac{1}{2} \sum_{i=1}^{n} \left[ \frac{(y_i - \eta(\xi_i, \Theta))^2}{\sigma^2(y_i)} + \frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} \right].$$

(2)

In functional (2), the values $\xi_i$ are unknown, but only their confidence intervals are known. The values $\xi_i$ are defined from the condition:

$$\frac{\partial F}{\partial \xi_i} \bigg|_{\xi_i = \hat{\xi}_i} = 0, \quad i = 1, n.$$

(3)

Parameter estimates $\hat{\Theta}$ are from the condition:

$$\frac{\partial F}{\partial \theta_b} \bigg|_{\theta_b = \hat{\theta}_b} = 0; \quad b = 1, S.$$

(3)

When defining $\xi_i$ one must ensure that the following type of constraint is enforced:

$$|x_i - \xi_i| \leq k\sigma(x_i),$$

where $k$ – the coefficient determined based on the condition of trust.

Thus, the problem of minimization of functional (2) under condition (3) is equivalent to solving a system of equations:

$$\sum_{i=1}^{n} \frac{y_i - \eta(\xi_i, \Theta)}{\sigma^2(y_i)} \cdot \frac{\partial \eta(\xi_i, \Theta)}{\partial \theta_b} = 0, \quad b = 1, S,$$

(4)
with \( \frac{x_i - \xi_i}{\sigma^2(x_i)} + \frac{y_i - \eta(\xi_i, \Theta)}{\sigma^2(y_i)} \cdot \frac{\partial \eta(\xi_i, \Theta)}{\partial \xi_i} = 0, \quad i = 1, n \).

When solving the system of equations (4) taking into account the condition (3), we obtain an iterative process that ends when one of the following conditions is met:

1) at the next step, the value of the functional (2) is less than the specified number \( \gamma \);

2) on neighboring iterations, the value of the functional (2) and the values of the parameter estimates \( \Theta \) differ slightly

\[
\left| \frac{F_v - F_{v+1}}{F_v} \right| \leq \gamma_1, \quad \max \left| \frac{\theta_b^v - \theta_b^{v+1}}{\theta_b^v} \right| \leq \gamma_2, \quad b = 1, S.
\]

where \( \gamma_1 \) and \( \gamma_2 \) – specified number;

3) iteration limit reached.

Variance estimates \( D(\hat{\Theta}) \) are determined from the matrix inverse of the matrix \( M \). Elements \( M_{be} \) are calculated by the formula:

\[
M_{be} = -\left. \frac{\partial^2 F}{\partial \theta_b \partial \theta_e} \right|_{\Theta = \hat{\Theta}}, \quad b, t = 1, S.
\]

Parameter estimates \( \hat{\Theta} \) are determined based on experimental values of features containing random errors. The values of the parameter estimates \( \hat{\Theta} \) in each particular experiment may differ from the values of the parameters \( \Theta \) and, therefore, there is still a certain amount of uncertainty [7]. The magnitude of this uncertainty can be found from the variances of the parameters \( D(\hat{\Theta}) \). Thus it is possible to tell that with a certain share of probability \( \Theta \) lies in an interval \( \hat{\Theta} \pm \sqrt{D(\hat{\Theta})} \).

The values of the functions \( \eta_j(x, \Theta) \), at each point \( x = \xi_i \), also have some uncertainty \( \eta_j(x, \Theta) \pm \Delta \eta_j(x, \Theta) \), i.e. change within [8]:

\[
\eta_{\beta j}(x, \Theta) \leq \eta_j(x, \Theta) \leq \eta_{\beta j}(x, \Theta).
\]

It should be noted that since it \( \eta_j(x, \Theta) \) is the product of the conditional density of the probability distribution and the a priori probability \( A_j \) for the event, the condition must be satisfied:

\[
\int \sum_{j=1}^{M} \eta_{\beta j}(x, \Theta)dx = \int \sum_{j=1}^{M} \eta_{\beta j}(x, \Theta)dx = 1.
\]

In addition to point estimates, for all values \( x = \xi_i \), we define interval estimates of functions \( \eta_j(x, \Theta) \). This can be done by knowing the unbiased parameter estimates \( \hat{\Theta} \) and parameter variances \( D(\hat{\Theta}) \) by the following formula:

\[
P\left( \eta_j(x, \Theta) - t_\gamma \sqrt{D(\eta_j(x, \Theta))} \leq \eta_j(x, \Theta) \leq \eta_j(x, \Theta) + t_\gamma \sqrt{D(\eta_j(x, \Theta))} \right) = \gamma, \quad (5)
\]

where \( \gamma \) – confidence probability, \( t_\gamma \) – quantile of student's distribution, \( D(\eta_j(x, \Theta)) \) – the variance of the function value \( \eta_j(x, \Theta) \), which in the case of uncorrelated \( S \)-parameters can be calculated by the formula:
\[ D(\eta_j(x, \Theta)) = \sum_{k=1}^{s} \left( \frac{\partial \eta_j(x, \Theta)}{\partial \theta_k} \right)^2 |_{\Theta = \hat{\Theta}} D(\theta_k). \]

3. Average risk assessment

Average risk can be considered as one of the criteria for the effectiveness of Bayesian Belief Network. Since random errors are unavoidable in real-world Bayesian Belief Network, taking them into account when determining the average risk increases the efficiency of decision-making. We determine the value of the average risk and its interval estimates, taking into account the uncertainty of the initial data.

Denote \( a_i, \mathcal{T} = 1, p \) - possible actions that can be done. Then, if a particular value is obtained as a result of the observation \( x \) and an action \( a_i \) is going to be performed, then the losses \( A_j \) are incurred for the event \( L(a_i|A_j) \).

Action-related losses \( a_i \), or a posteriori risk, are determined by the formula:

\[ r(a_i|x) = \sum_{j=1}^{M} L(a_i|A_j) P(A_j|x). \]  

Let be \( a(x) \) – the deciding function that tells which action of \( a_i, \mathcal{T} = 1, p \), should be taken for any observable value \( x \).

The average risk, i.e. losses associated with a certain decision rule, has the form:

\[ R = \int_x r(a(x)|x)P(x)dx, \]  

or taking into account (6) we get:

\[ R = \int_x P(x) \sum_{j=1}^{M} L(a(x)|A_j) P(A_j|x)dx. \]

In accordance with the Bayes formula, taking into account the previously accepted designations and for all \( x, \xi \), the average risk formula will take the form:

\[ R = \int_x \sum_{j=1}^{M} L(a(x)|A_j) \eta_j(x, \Theta)dx. \]  

In addition to point estimates, as shown above, interval estimates can be obtained for functions \( \eta_j(x, \Theta) \) using formula (5). Taking into account the interval estimates of functions, the value of the average risk will vary within some limits \( R_{\gamma} \leq R \leq R_{\gamma} \), i.e., for the average risk, interval estimates should be determined. The upper and lower bounds of the average risk with a given probability \( \gamma \) can be obtained by the formulas:

\[ R_{\gamma} = \int_x \sum_{j=1}^{M} L(a(x)|A_j) \eta_{j\gamma}(x, \Theta)dx, \]

\[ R_{\gamma} = \int_x \sum_{j=1}^{M} L(a(x)|A_j) \eta_{j\gamma}(x, \Theta)dx, \]

where \( \eta_{j\gamma}(x, \Theta) \) and \( \eta_{j\gamma}(x, \Theta) \) are determined from (5).

4. Modeling of bayesian belief network

Consider an example of a simple Bayesian Belief Network having two possible states \( A_1 \) and \( A_2 \). The parametric form of \( P(x|A_j) \) for each event is known and the a priori probabilities of events are known \( P(A_j) \).
It is known that conditional probability densities \( P(x|A_j) \) correspond to the normal law with parameters \((\mu_j, \sigma^2_j)\) and a priori probabilities are equal \( P(A_j) = P(A_2) = 0.5 \)

\[
P(x|A_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x-\mu_j)^2}{2\sigma^2} \right].
\]

The initial data are:
1) estimates of the parameters of functions \( \eta_j(x, \Theta) \) obtained by the least squares method;
2) values of observed quantities \( y_i \) and \( x_i \);
3) the measurement errors of the observed quantities \( y_i \) and \( x_i \), and the error values for each \( i \) – measurement are equal

\[
\sigma(y) = \sigma(y_i) = 0.3, \quad \sigma(x) = \sigma(x_i) = 40, \quad i = \overline{1,n}.
\]

Knowing the values \( y_i, x_i \) and their errors, the methods of confluent analysis determine unbiased estimates of the parameters. Based on (4) and taking into account that the conditional probability densities correspond to the normal law, we obtain a system of nonlinear equations:

\[
\sum_{i=1}^{n} \frac{y_i - \eta(\xi_i, \Theta)}{\sigma^2(y)} \eta(\xi_i, \Theta) = 0,
\]

\[
\sum_{i=1}^{n} \frac{y_i - \eta(\xi_i, \Theta)}{\sigma^2(y)} \frac{\eta(\xi_i, \Theta)}{\sigma^2(y)} = 0,
\]

provided:

\[
\frac{x_i - \xi_i}{\sigma^2(x_i)} \frac{y_i - \eta(\xi_i, \Theta)}{\sigma^2(y_i)} = 0, \quad i = \overline{1,n}.
\]

This system of equations can be solved by one of the iterative methods, for example, the Gauss method. At the same time, the parameter estimates obtained by the least squares method are used as zero approximations for estimating parameters by confluent analysis methods. The matrix \( M \) is used to determine the variance of parameter estimates.

Table 1 shows the values of parameter estimates for the two states \( A_1 \) and \( A_2 \), obtained by the least squares method and the methods of confluent analysis.

|          | Least square method |                     | Methods of confluent analysis |                     |
|----------|---------------------|---------------------|-----------------------------|---------------------|
| \( \mu_1 \) | 400                 | \( \sigma_1 \) \[100\] | 406±26                      | \( \mu_2 \) \[600\] | 104±4.5         | 585±32          | 79±2.9          |

According to the parameter estimates, we determine the point estimates of the functions and by the formula (5) with a probability of 0.95 we find the interval estimates of the functions \( \eta_j(x, \Theta) \). At the same time for \( n = 12, t_{p} = 2.2 \). The variance of the function value \( \eta_j(x, \Theta) \) is determined by the formula:
To determine the average risk, one must choose a decisive rule. Let us use the Bayesian decision rule and assume that the losses $L(a_t|A_j) = 0$ at $t = j$ and $L(a_t|A_j) = 1$ at $t \neq j$, i.e., there are no losses when the correct solution is chosen and are single for any decision-making errors.

Areas corresponding to different solutions are separated by boundaries. For two possible states $A_i$ and $A_2$, there will be two solution areas. For a Bayesian solving rule the separating boundary of the solution domains for all $x = \xi$ is defined from the equation:

$$P(x|A_i)P(A_i) = P(x|A_2)P(A_2).$$

The dividing boundary, taking into account that $P(x|A_j)$ correspond to the normal law and the a priori probabilities are equal, has the form:

$$x_0 = \frac{\mu_2 - \mu_1}{\sigma_2^2 - \sigma_1^2} \pm \frac{\sigma_1}{\sigma_2} \sqrt{\frac{(\mu_2 - \mu_1)^2 + (\sigma_2^2 - \sigma_1^2) \ln(\sigma_2^2/\sigma_1^2)}{\sigma_2^2 - \sigma_1^2}}. $$

When solving the problem by methods of confluent analysis we obtain $x_0 = 495.7$.

Table 1 shows that when solving the problem by the least squares method, the values $\sigma_j$ are equal. In this case the dividing border will take a simpler form:

$$x_0 = \frac{\mu_1 + \mu_2}{2} = 500.$$

For this problem $\mu_1 < \mu_2$, therefore, of (7) for all $x = \xi$, the average risk is determined by the formula:

$$R = \int_{-\infty}^{x_0} \eta_2(x, \Theta)dx + \int_{x_0}^{+\infty} \eta_1(x, \Theta)dx.$$

Find the value of the average risk in two ways. The parameter estimates obtained by the method of least squares $R = 0.08$, the parameter estimates obtained by the methods of confluent analysis $R_{ca} = 0.1$. The analysis of the results shows that the values of the average risk determined in two ways do not coincide and taking into account the random nature of the initial data leads to an increase in the value of the average risk.

Taking into account the random nature of the initial data, interval estimates of the average risk are determined by the formulas:

$$R_{II} = \int_{-\infty}^{x_0} \eta_{2II}(x, \Theta)dx + \int_{x_0}^{+\infty} \eta_{1II}(x, \Theta)dx, \quad R_{B} = \int_{-\infty}^{x_0} \eta_{2B}(x, \Theta)dx + \int_{x_0}^{+\infty} \eta_{1B}(x, \Theta)dx.$$

where $\eta_{2B}(x, \Theta)$ and $\eta_{1B}(x, \Theta)$ are determined from (5), taking into account the formula (8).

With a probability of 0.95 the value of the average risk is in the range $0.08 < R_{ca} < 0.12$.

5. Conclusion

Traditional methods of estimation of parameters of functional dependences allow receiving only point estimations of parameters and do not completely consider errors of initial data. The use of confluent
analysis methods makes it possible to take into account the uncertainty of the initial information and obtain unbiased estimates of parameters and their variance. Point and interval estimates of functional dependences are found by parameter values.

The proposed method of accounting for the uncertainty of the initial information by the methods of confluent analysis and interval estimation of conditional probability densities allows increasing the reliability of decision-making in Bayesian confidence networks.

Acknowledgments
This work was supported by the Russian Foundation for Basic Research, project no. 19-03-00271, and by the Russian Foundation for Basic Research and the Government of Kaluga District, project no. 18-41-400001.

References
[1] Pourret O Bayesian networks: a practical guide to applications Wiley 2008
[2] Fuller W A Measurement error models Wiley 1987
[3] Gleser L J Statist. Anal. Meas. Error Models and Appl. 1989 99
[4] Gustafson P Measurement error and misclassification in statistics and epidemiology 2003
[5] Greshilov A A Mathematical methods of decision making 2014
[6] Gagarin Yu E, Lapshinova E N, Petrov V I and Stepovich M A 2007 Journal of Surface Investigation. X-ray, Synchrotron and Neutron Techniques 10 31
[7] Gagarin Yu E 2019 Modeling, Optimization and Information Technology 7
[8] Gagarin Yu E and Gagarina S N 2018 Vestnik Universiteta 12 64