Probabilistic Constrained Optimization: Methodology and Applications (S. P. Uryasev, Editor), pp. 102-131

Optimization in the Space of Distribution Functions and Applications in the Bayes Analysis [1]

Alexandr N. Golodnikov (golodnik@d130.icyb.kiev.ua)
V.M.Glushkov Institute of Cybernetics
Prospect Glushkova,40, Kiev 252022, Ukraine

Pavel S. Knopov (knopov@d130.icyb.kiev.ua)
V.M.Glushkov Institute of Cybernetics
Prospect Glushkova,40, Kiev 252022, Ukraine

Panos M. Pardalos (pardalos@ufl.edu)
Center for Applied Optimization and ISE Department,
University of Florida, Gainesville, FL 3261-6595, USA

Stanislav P. Uryasev (uryasev@ise.ufl.edu)
ISE Department,
University of Florida, Gainesville, FL 3261-6595, USA

Abstract

This paper is stimulated by reliability estimation problems for safety systems of Nuclear Power Plants. A new approach for calculating robust Bayes estimators is considered. Upper and lower bounds for Bayes estimates, provided that a prior distribution satisfies available prior information, are constructed. The problems of calculating lower and upper bounds for Bayes estimates is reduced to optimization on a set of distributions satisfying available prior information. It was demonstrated that Bayes estimates of parameters are sensitive to the type of a prior distribution function. Analysis of the reliability data of a nuclear safety system was conducted using the developed methodology. The robust estimates were compared to Bayes estimates traditionally used in nuclear industry.

[1] Research was supported in part by a NATO grant (SST.CLG 975032)

1 Introduction

Analysis of nuclear safety data, as well as many other applications, rely on using the Bayes estimation techniques. For instance, Probabilistic Safety Assessments (PSAs) depend upon data to quantify the basic failure-related events and the initiating events frequencies. Plant-specific data are derived from plant records and are processed to estimate reliability parameters for use in PSA. Statistical techniques that are usually used for estimation of reliability parameters are based on sampling theory methods. However, sampling theory methods are inappropriate for those Nuclear Power Plant (NPP) components that have scarce data samples. Although there are plenty of historical operating data for other NPPs, these data cannot be incorporated in reliability parameter estimates within the sampling theory methods. In order to overcome this shortcoming of the sampling theory, the Bayesian approach can be used which requires less sample data to achieve the same quality of inferences than the methods based on sampling theory. The Bayesian methods permit the combining of plant-specific raw data with other relevant information available from reliability studies.

For a recent review of literature and collection of papers in the area of robust Bayesian estimates see Springer Verlag volume [5]. This volume includes state-of-the-art results, such as the paper on Linearization techniques in Bayesian Robustness [10]. On application of the optimization approaches in statistics see paper [11].

A major criticism of the Bayesian approach is related to the ability of the investigator to select a true distribution function. The justification of a prior distribution frequently is a practical difficulty in the application of the Bayesian approach. This shortcoming of the Bayesian approach is usually overcome by using the Empirical Bayesian decision procedures. The Empirical Bayesian decision procedure is an efficient tool for combining either existing sets of reliability data or reliability parameter estimates from various sources. One of the advantages of such methods is their asymptotic optimality. However, the practical rate of convergence of the Empirical Bayesian risk to the minimum Bayesian risk can be quite slow. Therefore, in cases when small datasets of reliability data or reliability parameter estimates are available, the accuracy of the approximation of the Bayesian estimator by the Empirical Bayesian estimators is never really known. In such cases, it is desirable to estimate how far the calculated Bayesian estimate is from the true Bayesian estimate.

Similar problems take place in calculating two-sided Bayes probability intervals, which are used in uncertainty evaluation. Since this requires knowledge of the true posterior distribution, which depends on the true prior distribution, there can be situations where the two-sided Bayes probability interval derived using the assumed prior distribution differs from that derived using the true prior distribution.

Since in practice the true prior distribution is never known, it is reasonable to consider that any distribution function that corresponds to the same available prior information has the same justification to be used as a prior distribution. Using the beta distribution as a prior in binomial sampling and the gamma distribution as a prior in Poisson sampling is often justified by their mathematical tractability and versatility. However, this reasoning is not so important now because of availability of high-speed computational resources.

Therefore, when prior information is insufficient to accurately specify a prior distribution, the most attractive Bayes estimate is a robust Bayes estimate derived under an assumption that the true prior distribution corresponds to the available prior information.

In this connection, the following problems have been addressed: (1) development of methods of calculation of upper and lower bounds for Bayes estimates provided that a prior distribution satisfies available prior information; (2) development of methods of calculation of robust Bayes estimates; and (3) development of methods of calculation of a robust two-sided Bayes probability interval provided that a prior distribution satisfies available prior information.

This paper presents mathematical statements of the problems mentioned above, numerical methods for their solving, example of failure data analysis, and comparison of obtained results with traditionally used Bayes estimates. This paper is based on the approach developed in [4, 3].

2 Sensitivity to a prior distribution in a binomial sampling

Let us consider the following slightly modified example, borrowed from [9]. The WASH-1400 "Reactor Safety Study" (1975) reported the following data on the number of pump failures observed in 1972 in eight pressurized water reactors (PWRs) in commercial operation in the United States.

Table 1: Data on the number of pumps failures.

j	n_j	s_j	x_j	$t_j(h)$	$\frac{x_j}{n_j}$	$\frac{s_j}{n_j}$
1	50	12	38	438	0.76	0.24
2	50	2	48	438	0.96	0.04
3	50	1	49	438	0.98	0.02
4	50	5	45	438	0.9	0.1
5	50	6	44	438	0.88	0.12
6	50	0	50	438	1	0
7	50	1	49	438	0.98	0.02
8	50	3	47	438	0.94	0.06
Total:	400	30	370	3504	7.4	0.6

Here s_j is the observed number of failures in n_j pumps, and t_j is the total test time in hours. The data in rows of the table 1 can be treated as results in series of $N\,(N = 8)$ repeated trials.

The situation where n is fixed in advance, and the number of failures is left to chance, is known as binomial sampling. The probability distribution of the number of failures is the binomial distribution given by

$$Pr\{s\,\text{failures will occur in } n\,\text{trials}\} = f(s|p) = \frac{n!}{(n-s)!s!}p^s(1-p)^{n-s}, \qquad (1)$$

where $s = 0, 1, \ldots, n$, $0 < p < 1$; p is a constant probability of failure in each trial. In the sampling theory approach, the maximum likelihood estimator of unknown parameter p in jth series of repeated trials is

$$\hat{p}_j = \frac{s_j}{n_j}. \qquad (2)$$

In the Bayesian approach, the parameter p is assumed to be a random value and $p_1, p_2, \ldots, p_N$ are its statistically independent realization for the sequence of trials. The most widely used prior distribution for p is the beta distribution $\mathcal{B}(s_0, n_0)$ with the following probability density function

$$g(r) = \begin{cases} \frac{\Gamma(n_0)}{\Gamma(s_0)\Gamma(n_0-s_0)} r^{s_0-1}(1-r)^{n_0-s_0-1}, & 0 \le r \le 1 \\ 0, & \text{otherwise,} \end{cases}$$

where $\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx$ is the gamma function. Using the beta distribution as a prior is often justified by its mathematical tractability and versatility.

Let us consider a family of distributions defined on [0,1] with the same mean μ and variance σ^2. First, let us express the mean and variance of a prior distribution in terms of sample moments by using the method of moments [9]. The conditional mean and variance of $\hat{p}_j$, conditioned upon the unknown value of p_j in jth trial, are

$$E(\hat{p}_j|p_j) = p_j \qquad (3)$$

$$Var(\hat{p}_j|p_j) = \frac{p_j(1-p_j)}{n_j}. \qquad (4)$$

The unconditional expectation of $\hat{p}_j$ is

$$E(\hat{p}_j) = E_{p_j}[E_{\hat{p}_j|p_j}(\hat{p}_j|p_j)] = E_{p_j}[p_j] = \mu. \qquad (5)$$

The relation between unconditional and conditional variance is given by

$$Var(\hat{p}_j) = E_{p_j}[Var_{\hat{p}_j|p_j}(\hat{p}_j|p_j)] + Var_{p_j}[E_{\hat{p}_j|p_j}(\hat{p}_j|p_j)].$$

By substituting (3) and (4) in the previous formula we have

$$\begin{aligned} Var(\hat{p}_j) &= E\left[\frac{p_j(1-p_j)}{n_j}\right] + Var(p_j) = \\ &= \frac{1}{n_j}E(p_j) - \frac{1}{n_j}E(p_j^2) + Var(p_j) = \\ &= \frac{1}{n_j}E(p_j) - \frac{1}{n_j}[Var(p_j) + E^2(p_j)] + Var(p_j). \end{aligned}$$

Therefore,

$$Var(\hat{p}_j) = \frac{1}{n_j}\mu - \frac{1}{n_j}[\sigma^2 + \mu^2] + \sigma^2. \tag{6}$$

The method of moments equates the sample moments to their expected values and solves for the μ and σ^2. The first weighted sample moment is

$$\bar{p}_w = \frac{\sum_{j=1}^{N} n_j \hat{p}_j}{\sum_{j=1}^{N} n_j} = \frac{1}{N^*}\sum_{j=1}^{N} s_j, \quad N^* = \sum_{j=1}^{N} n_j. \tag{7}$$

And the second weighted sample moment is

$$m_w^2 = \frac{1}{N^*}\sum_{j=1}^{N} n_j \hat{p}_j^2. \tag{8}$$

The expected value of the first weighted sample moment is

$$E\bar{p}_w = \frac{1}{N^*}\sum_{j=1}^{N} Es_j = \frac{\mu}{N^*}\sum_{j=1}^{N} n_j = \mu. \tag{9}$$

The expected value of the second weighted sample moment is

$$\begin{aligned} E(m_w^2) &= \frac{1}{N^*}\sum_{j=1}^{N} n_j E(\hat{p}_j^2) = \\ &= \frac{1}{N^*}\sum_{j=1}^{N} n_j [Var(\hat{p}_j) + E^2(\hat{p}_j)] \end{aligned} \tag{10}$$

By substituting (5) and (6) in (10) we have

$$E(m_w^2) = \frac{1}{N^*}\sum_{j=1}^{N}(\mu - [\sigma^2 + \mu^2] + n_j[\sigma^2 + \mu^2]) = \frac{N}{N^*}(\mu - [\sigma^2 + \mu^2]) + \sigma^2 + \mu^2. \tag{11}$$

By equating (9) and (11) to $\bar{p}_w$ and m_w^2, respectively and solving for μ and σ^2, we obtain

$$\mu = \bar{p}_w \tag{12}$$

and

$$\sigma^2 = \frac{N^*(m_w^2 - \bar{p}_w^2) - N(\bar{p}_w - \bar{p}_w^2]}{N^* - N}. \tag{13}$$

The first weighted sample moment of the prior distribution for data from the Table 1 is $\bar{p}_w = 0.075$ and the second weighted sample moment is $m_w^2 = 0.011$. Therefore, estimates of mean and variance of the prior distribution of the parameter p obtained by using (12), (13), and data from Table 1 equal

$$\mu = 0.075, \quad \sigma^2 = 0.004069. \tag{14}$$

2.1 Beta distribution is used as a prior.

Taking into consideration that the mean and variance of the beta distribution $\mathcal{B}(s_0, n_0)$ are

$$\mu = \frac{s_0}{n_0}$$

and

$$\sigma^2 = \frac{s_0(n_0 - s_0)}{n_0^2(n_0 + 1)},$$

and solving for s_0 and n_0, we obtain

$$n_0 = \frac{\mu(1-\mu)}{\sigma^2} - 1 = 16.05016\,,$$

$$s_0 = \mu n_0 = 1.203762\,.$$

According to [9], the Bayesian point estimator for p under a squared-error loss function and beta prior distribution is

$$\tilde{p}_{\mathcal{B}} = E(p|s) = \frac{s + s_0}{n + n_0}, \tag{15}$$

2.2 Step-function with 3 steps is used as a prior.

Suppose that the true prior is one of the following step functions that have steps at three points:

$$G_0(p) = \begin{cases} 0, & \text{if} \;\; 0 \le p < 0.07, \\ 0.396773, & \text{if} \;\; 0.07 \le p < 0.071, \\ 0.995267, & \text{if} \;\; 0.071 \le p < 1, \\ 1, & \text{if} \;\; p = 1, \end{cases} \tag{16}$$

$$G_1(p) = \begin{cases} 0, & \text{if} \;\; 0 \le p < 0.001, \\ 0.120358, & \text{if} \;\; 0.001 \le p < 0.081, \\ 0.996052, & \text{if} \;\; 0.081 \le p < 1, \\ 1, & \text{if} \;\; p = 1, \end{cases} \tag{17}$$

$$G_2(p) = \begin{cases} 0, & \text{if} \;\; 0 \le p < 0.001, \\ 0.426296, & \text{if} \;\; 0.001 \le p < 0.129, \\ 0.434103, & \text{if} \;\; 0.129 \le p < 0.13, \\ 1, & \text{if} \;\; 0.13 \le p \le 1, \end{cases} \tag{18}$$

$$G_3(p) = \begin{cases} 0, & \text{if} \;\; 0 \le p < 0.003, \\ 0.439741, & \text{if} \;\; 0.003 \le p < 0.131, \\ 0.713175, & \text{if} \;\; 0.131 \le p < 0.132, \\ 1, & \text{if} \;\; 0.132 \le p \le 1, \end{cases} \tag{19}$$

It can be verified that each of these distribution functions has the same mean $\mu = 0.075$ and variance $\sigma^2 = 0.004069$.

For each of these distributions, Bayes estimates were calculated for number of failures $s = 0, 1, 2, 3$. Results obtained are compared with the corresponding Bayes estimates derived under an assumption that beta distribution is a true prior (see Table 2.) In the Table 2 each column corresponds to Bayes estimates calculated under an assumption that the true prior is a distribution function exhibited in the first row.

Table 2: Bayes estimates calculated for different prior distributions

s	$BETA$	$G_0(p)$	$G_1(p)$	$G_2(p)$	$G_3(p)$
0	0.018225	0.070588	0.00906024	0.001173	0.003165
1	0.033365	0.070592	0.07363717	0.022494	0.010802
2	0.048505	0.070596	0.08090801	0.125804	0.101286
3	0.063645	0.070599	0.08099895	0.129957	0.130724

Data in the Table 2 demonstrate that Bayes estimates are sensitive to the type of a prior distribution function that has the same first two moments.

3 Problem statement

As a rule, in PSA applications, sources of prior information are plant-specific data that are derived from plant records; observational data from previous comparable experiments; and expert estimates. Important sources of prior information are reports of US Nuclear Regulatory Commission (NUREGs) and generic data books where parameters of a prior distribution such as the mean, the 5th, 50th and 95th percentile values are presented. In practical applications, the range of possible values of an estimated parameter can be restricted by a finite interval $[a, b]$, and prior data from both objective and subjective sources can be expressed in the form of equalities and inequalities of the following types

$$\int_a^b dH(\omega) = 1, \tag{20}$$

$$\int_a^b f_i(\omega) dH(\omega) \leq d_i, \tag{21}$$

$$i = 1, 2, \ldots, m,$$

or

$$\int_a^b dH(\omega) = 1, \tag{22}$$

$$\int_a^b \omega^i dH(\omega) = \mu_i, \tag{23}$$

$$i = 1, \ldots, m,$$

where functions $f_i(\omega)$, $i = 1, 2, \ldots, m$, are not necessarily continuous. Let K denote one of the classes of distribution functions defined by (20)-(21) or (22)-(23). By means of expressions presented above, we can describe sufficiently general prior information about prior distributions. For example, they can be used to define a class of distribution functions with fixed first m moments, restrictions on q quantiles or their support. Obviously, in the later two cases, functions $f_i(\omega), \quad i = 1, 2, \ldots, m$, are not continuous.

Given sample data x, the Bayes estimate δ according to Bayesian procedures can be found by minimizing a posterior risk

$$\min_\delta \int_a^b L(\omega, \delta) dG(\omega|x) = \min_\delta \frac{\int_a^b L(\omega, \delta) f(x|\omega) dH(\omega)}{}, \int_a^b f(x|\omega) dH(\omega) \tag{24}$$

where $f(x|\omega)$ is the conditional probability density function of x given ω, which describes the sampling model; $G(\omega|x)$ is the posterior distribution of ω given x, which is calculated by using Bayes' theorem; and $L(\omega, \delta)$ is a loss function for given estimate δ.

For the squared-error loss function $L(\omega, \delta)$, the Bayes estimate is the posterior mean of ω given x

$$\delta = \frac{\int_a^b \omega f(x|\omega) dH(\omega)}{\int_a^b f(x|\omega) dH(\omega)} . \tag{25}$$

As it was demonstrated in the previous section, for given sample data x, there are ranges of values of the first two moments where Bayes estimates are sensitive to the type of the prior distribution function that has the same first two moments. Since any distribution function that corresponds to the same available prior information has the same justification to be used as a prior distribution, it is important to find a range of Bayes estimates that can be derived for all distributions from class K. Its upper, c^*, and lower, c_*, bounds can be used to determine how far the calculated Bayes estimate is from the true Bayes estimate. The range of possible values of Bayes estimates, $[c_*, c^*]$, can be treated as a measure of sensitivity of Bayes estimates with respect to selection of a prior distribution under available prior information. Since points of the segment $[c_*, c^*]$ are derived under the same prior information, any of them can be used as a Bayes estimate in PSA applications. According to the conservative approach, it is reasonable to use the largest value, c^*, of the range of possible Bayes estimates in PSA applications. This value is a conservative Bayes estimate because any other Bayes estimates derived under another distribution satisfying the same prior information can only reduce the estimate of a failure probability and thereby improve PSA estimates.

In order to find such bounds for given sample data x, it is necessary to solve the following optimization problems

$$\frac{\int_a^b \omega f(x|\omega) dH(\omega)}{\int_a^b f(x|\omega) dH(\omega)} \rightarrow min, \tag{26}$$

subject to (20) - (21), or (22) - (23), and

$$\frac{\int_a^b \omega f(x|\omega)dH(\omega)}{\int_a^b f(x|\omega)dH(\omega)} \to max, \tag{27}$$

subject to (20) - (21), or (22) - (23).

The reasons presented above change the point of view on estimating the Bayes probability interval, which is used in uncertainty evaluation. According to the Bayesian approach, once the posterior distribution of ω given x has been obtained, a symmetric $100(1-\gamma)\%$ Bayes probability interval estimate of ω is obtained by solving the two equations

$$\frac{\int_a^{\omega_*} f(x|\omega)dH(\omega)}{\int_a^b f(x|\omega)dH(\omega)} = \frac{\gamma}{2}, \tag{28}$$

and

$$\frac{\int_{\omega^*}^b f(x|\omega)dH(\omega)}{\int_a^b f(x|\omega)dH(\omega)} = \frac{\gamma}{2}, \tag{29}$$

for the lower limit ω_* and the upper limit ω^*, so that $Pr(\omega_* \leq \omega \leq \omega^*|x) = 1-\gamma$.

If there is no accurate information about the true prior distribution, determining the Bayes probability interval by using equations (28) - (28) for any prior distribution from the class K will no longer guarantee that the true unknown parameter belongs to it with probability $1-\gamma$. In this case, the new procedure should be applied to determine a 'generalized' Bayes probability interval, which would guarantee that the true unknown parameter belongs to it with probability at least $1-\gamma$.

For this reason, if the only prior information available is that the prior distribution belongs to the class K, a symmetric $100(1-\gamma)\%$ Bayes probability interval estimate of ω can be obtained by solving the following two problems.

$$\omega_* \to \sup \tag{30}$$

subject to

$$\frac{\int_a^{\omega_*} f(x|\omega)dH(\omega)}{\int_a^b f(x|\omega)dH(\omega)} \leq \frac{\gamma}{2}, \tag{31}$$

for all $H(\omega) \in K$, and

$$\omega^* \to \inf \tag{32}$$

subject to

$$\frac{\int_{\omega^*}^b f(x|\omega)dH(\omega)}{\int_a^b f(x|\omega)dH(\omega)} \leq \frac{\gamma}{2}, \tag{33}$$

for all $H(\omega) \in K$.

The problems (30)-(31) and (32)-(33) can be expressed in the following equivalent forms

$$\omega_* \to \sup \tag{34}$$

subject to

$$\sup_{H(\omega)\in K} \frac{\int_a^{\omega_*} f(x|\omega)dH(\omega)}{\int_a^b f(x|\omega)dH(\omega)} \le \frac{\gamma}{2}, \tag{35}$$

and

$$\omega^* \to \inf \tag{36}$$

subject to

$$\inf_{H(\omega)\in K} \frac{\int_a^{\omega^*} f(x|\omega)dH(\omega)}{\int_a^b f(x|\omega)dH(\omega)} \ge 1 - \frac{\gamma}{2}. \tag{37}$$

Another definition of the sensitivity of the Bayes estimator can be given in terms of the Bayes risk. The measure of sensitivity of Bayes estimator $\delta_0(x)$ derived under an assumed prior distribution $H_0(\omega) \in K$ to the type of the prior distributions $H(\omega) \in K$ that correspond to the same available prior information can be expressed as follows:

$$\chi(\delta_0(x)) = \max_{H(\omega)\in K} \frac{r_H(\delta_0(x)) - r_{H_0}(\delta_0(x))}{r_{H_0}(\delta_0(x))}, \tag{38}$$

where

$r_{H_0}(\delta_0(x))$ is the Bayes risk that corresponds to the Bayes estimator $\delta_0(x)$ provided that distribution $H_0(\omega)$ is the true prior;

$r_H(\delta_0(x))$ is the Bayes risk that corresponds to the Bayes estimator $\delta_0(x)$ provided that distribution $H(\omega) \in K$ is the true prior.

In order to calculate the value of $\chi(\delta_0(x))$, it is necessary to solve the following problem

$$\max_{H(\omega)\in K} r_H(\delta_0(x)) = \max_{H(\omega)\in K} \int_a^b \int_X L(\omega, \delta_0(x)) f(x|\omega) dx dH(\omega),$$

where

$L(\omega, \delta_0(x))$ is a loss function for given decision function (estimator) $\delta_0(x)$;

$f(x|\omega)$ is the conditional probability density function (sampling model). Let

$$f_0(\omega) = \int_X L(\omega, \delta_0(x)) f(x|\omega) dx \tag{39}$$

Then, the problem of calculating the sensitivity of a given Bayes estimator $\delta_0(x)$ is reduced to the following problem of optimization of a linear functional over prior distributions from the class K of distribution functions described by available prior information.

$$\int_a^b f_0(\omega) dH(\omega) \to \max, \tag{40}$$

subject to (20) - (21), or (22) - (23).

It is also desirable to estimate how far the calculated Bayes risk derived under assumed prior distribution is from the true Bayes risk. The problems of calculating lower and upper bounds for Bayes risk are stated as the problem of minimization and maximization of functional (40) over distributions that satisfy (20) - (21), or (22) - (23).

This section considers stochastic programming problems where it is required to optimize linear or linear-fractional functionals over one-dimensional distribution functions that belong to some class, defined by restrictions (20) - (21), or (22) - (23). The next sections presents numerical methods for solving such problems for a general multi-dimensional case.

4 Optimization linear functional under constraints of inequality type

Consider the following optimization problem:

$$\varphi(H) = \int_X f_0(x) dH(x) \to \inf, \tag{41}$$

subject to

$$\int_X dH(x) = 1, \tag{42}$$

$$\psi_i(H) = \int_X f_i(x) dH(x) \leq a_i, \tag{43}$$

$$i = 1, 2, \ldots, m,$$

where X is a compact set from R^n.

Denote by $K_m(X)$ the set of distribution functions that satisfy (42)-(43). Let us consider two cases:

1. Functions $f_\nu(x)$ are lower semi-continuous, $\nu = 0, 1, \ldots, m$.
2. Functions $f_\nu(x)$ are upper semi-continuous, $\nu = 0, 1, \ldots, m$.

Case 1. Let functions $f_\nu(x)$ be lower semi-continuous, $\nu = 0, 1, \ldots, m$ and a_0 is a number such that for some distribution function $H(x)$, which satisfies (42)-(43), the following inequality holds

$$\varphi(H) = \int_X f_0(x) dH(x) \leq a_0. \tag{44}$$

Then the problem (41)-(43) can be written in the following form:

$$\varphi(H) = \int_X f_0(x) dH(x) \to \inf, \tag{45}$$

subject to

$$\int_X dH(x) = 1, \tag{46}$$

$$\int_X f_\nu(x) dH(x) \leq a_\nu, \tag{47}$$

$$\nu = 0, 1, \ldots, m.$$

Following [2], [7], and [8] we show that the functional (41) achieves its minimal value and the solution of the problem (41)-(43) is a step-function with at most $m+1$ steps.

Let

$$z_\nu = f_\nu(x), \qquad x \in X, \qquad \nu = 0, 1, \ldots, m,$$

Define

$$Z = \{z : z = (f_0(x), f_1(x), \ldots, f_m(x)), x \in X, f_\nu(x) \le a_\nu, \nu = 0, 1, \ldots, m\}. \tag{48}$$

The following theorem holds.

Theorem 1.

Assume that functions $f_\nu(x), \nu = 0, 1, \ldots, m$, are lower semi-continuous. Then the set Z is closed.

Proof. Let $z^k \to z^*$, where $z^k = (f_0(x^k), f_1(x^k), \ldots, f_m(x^k))$ and $z^k \in Z$. Since the set X is compact then from the sequence $\{x^k\}$ we can select a convergent subsequence. Without any loss of generality, assume that $x^k \to x^* \in X$. Since functions $f_\nu(x), \nu = 0, 1, \ldots, m$, are lower semi-continuous on the set X, then $\forall \varepsilon > 0\ \exists N$ such that $\forall k \ge N$ the following inequalities hold

$$f_\nu(x^*) - f_\nu(x^k) < \varepsilon, \qquad \nu = 0, 1, \ldots, m.$$

From these inequalities and (48), we obtain

$$f_\nu(x^*) = f_\nu(x^*) - f_\nu(x^k) + f_\nu(x^k) \le a_\nu + \varepsilon,$$

$$\nu = 0, 1, \ldots, m.$$

Since ε is arbitrary positive value, these inequalities imply that $z^* \in Z$ and therefore the set Z is closed, which proves the theorem. $\diamond$

Consider the convex hull of Z:

$$coZ = \{z : z = \sum_{k=1}^{r} p_k z^k, z^k \in Z, \sum_{k=1}^{r} p_k = 1, p_k \ge 0, k = 1, 2, \ldots, r\},$$

where r is an arbitrary positive integer value.

Since Z is a closed set, then coZ is a closed set as well. Denote by G a set of vectors $Q = (\varphi(H), \psi_1(H), \ldots, \psi_m(H))$ whose components satisfy (47) for any $H(x)$ such that $\int_X dH(x) = 1$. Since step-functions with a finite number of steps belong to the set of distribution functions, then $coZ \subset G$.

Any distribution function $H(x)$ can be approximated by step-functions with a sufficiently large number N of steps and $\int_X f_\nu(x) dH(x)$ can be approximated with sums $\sum_{k=1}^{N} f_\nu(x^k) p_k$, $\nu = 0, 1, \ldots, m$, which belong to coZ. Since coZ is a closed set then limits of these sums belong to coZ. Therefore $G = coZ$.

Now, the problem (41)-(43) can be written in the following form:

$$z_0 \to \min \tag{49}$$

subject to

$$z = (z_0, z_1, \ldots, z_m) \in coZ. \tag{50}$$

In (49) we substitute 'inf' by 'min' because the set coZ is closed and therefore $\min z_0$ is achieved on the set coZ. A solution of the problems (49)-(50) belongs to the boundary of coZ from $m+1$-dimensional space. According to the Caratheodory theorem, each point on the boundary of coZ can be represented as a convex combination of at most $m+1$ points from Z. Therefore the functional (41) achieves its minimal value and the solution of the problem (41)-(43) is a step-function with at most $m+1$ steps.

Assume that functions $f_\nu(x)$, $\nu = 0, 1, 2, \ldots, m$ have finite number of surfaces of discontinuity $A_1, A_2, \ldots, A_r$. Suppose that functions $f_\nu(x)$, $\nu = 0, 1, \ldots, m$ are continuous on $A_1, A_2, \ldots, A_r$, i.e. if A is the surface of discontinuity of function $f_\nu(x)$, then $\forall x' \in A$ and $\forall \varepsilon > 0$ $\exists \delta > 0$ such that $\forall x \in U_\delta(x') \bigcap A$ the following inequality holds

$$| f_\nu(x') - f_\nu(x) | < \varepsilon,$$

where $U_\delta(x') = \{x : \|x - x'\| < \delta\}$.

If $X \subset R^1$, a surface of discontinuity is a point of discontinuity.

Algorithm.

The idea of the algorithm is to reduce the initial infinite programming problem to a sequence of finite-dimensional problems. At each iteration we shall search for a solution of the problem (41)-(43) in a set of step-functions of finite support. The number of points in support sets increase from iteration to iteration.

On the s-th iteration construct partitions $R_s^i = (x_{1,i}^s, x_{2,i}^s, \ldots, x_{l_s,i}^s)$ of sets A_i, $i = 0, 1, \ldots, r$, such that $A_i \subset \bigcup_{j=1}^{l_s} U_{\lambda_s}(x_{j,i}^s)$, $R_s^i \subset R_{s+1}^i$, $i = 0, 1, \ldots, r$; $\lambda_s \to 0$ as $s \to \infty$, where $A_0 = X \setminus \bigcup_{i=1}^r A_i)$; $U_{\lambda_s}(x_{j,i}^s) = \{x : \|x - x_{j,i}^s\| < \lambda_s\}$.

Let $R_s = \bigcup_{i=0}^r R_s^i$, $R_s = (x_1^s, x_2^s, \ldots, x_{n_s}^s)$, where $n_s = (r+1) \cdot l_s$, $x_{(j+1)\cdot l_s + i}^s = x_{i,j}^s$, $i = 1, 2, \ldots, l_s$; $j = 0, 1, \ldots, r$.

Then, we solve the following linear programming problem:

$$\sum_{j=1}^{n_s} f_0(x_j) p_j \to \min \tag{51}$$

subject to

$$\sum_{j=1}^{n_s} p_j = 1, \tag{52}$$

$$\sum_{j=1}^{n_s} f_i(x_j) p_j \le a_i \qquad i = 1, 2, \ldots, m, \tag{53}$$

$$p_j \geq 0, \qquad x_j \in R_s, \qquad j = 1, 2, \ldots, n_s.$$

Let $p^s = (p_1^s, p_2^s, \ldots, p_{n_s}^s)$ be a solution of the problem (51)-(53). Since the number of basic variables in the problem (51)-(53) equals to $m+1$ then at most $m+1$ components of the vector p^s can be nonzero. Without any loss of generality, we can assume that the first $m+1$ components $p_1^s, p_2^s, \ldots, p_{m+1}^s$ are nonzero. Therefore, the solution of the problem (51)-(53), we can represent as

$$H_s(x) = (x_1^s, \ldots, x_{m+1}^s, p_1^s, \ldots, p_{m+1}^s), \tag{54}$$

where basic variable p_i^s corresponds to point x_i^s and appropriate basic column

$$(1, f_1(x_i^s), f_2(x_i^s), \ldots, f_m(x_i^s))'.$$

On the other hand, (54) can be considered as a representation of the distribution function, where the step-function $H_s(x)$ has step p_i^s at the point x_i^s, $i = 1, 2, \ldots, m$.

The solution $H_s(x)$ found at the sth iteration will be used as initial basic variables in solving the problem (51)-(53) at the $s+1$th iteration.

The following theorem holds.

Theorem 2.

Assume that

1. Functions $f_\nu(x)$ are lower semi-continuous and bounded on the set X, $\nu = 0, 1, 2, \ldots, m$.

2. X is a compact set from R^n.

3. $\exists \tilde{H}(x)$ such that the following restrictions hold

$$\int_X d\tilde{H}(x) = 1, \tag{55}$$

$$\psi_i(\tilde{H}) = \int_X f_i(x) d\tilde{H}(x) < a_i, \tag{56}$$
$$i = 1, 2, \ldots, m,$$

Then, limit of any convergent subsequence $\{H_{s_k}\}$ of the sequence $\{H_s\}$ belongs to the set X^*, where

$$X^* = \{H^*(x) : \varphi(H^*) = \min_{H(x) \in K_m(X)} \varphi(H)\}$$

and

$$\lim_{s \to \infty} \varphi(H_s) = \varphi(H^*).$$

Proof. As it was shown earlier, the solution of the problem (41)-(43) is a step-function $H^*(x)$ with at most $m+1$ steps. Suppose that $\varphi(H_s)$ does not converge to $\varphi(H^*)$. Then there exists an $\varepsilon_0 > 0$ and sequence of indexes $\{s_k\}$ such that the following inequality holds

$$\varphi(H_{s_k}) - \varphi(H^*) \geq \varepsilon_0 > 0 \tag{57}$$

Denote

$$H^*(x) = (x_1^*, \dots, x_{m+1}^*, p_1^*, \dots, p_{m+1}^*)$$

and

$$H_{s_k}(x) = (x_1^{s_k}, \dots, x_{m+1}^{s_k}, p_1^{s_k}, \dots, p_{m+1}^{s_k}),$$

where p_i^* and $p_i^{s_k}$ are the steps of the distribution functions $H^*(x)$ and $H_{s_k}(x)$ at the corresponding points x_i^* and $x_i^{s_k}$.

Since the point $Q = (\varphi(\tilde{H}), \psi_1(\tilde{H}), \dots, \psi_m(\tilde{H}))$ belongs to $coZ \subset R^m$ (not necessarily a boundary point of coZ), according to the Caratheodory theorem, it can be presented as a convex combination of at most $m+2$ points from Z. Therefore, there exists a step-function with at most $m+2$ steps that satisfies (55)-(56). Without any loss of generality we can assume that $\tilde{H}(x)$ is a step-function with at most $m+2$ steps. Denote

$$\tilde{H}(x) = (\tilde{x}_1, \dots, \tilde{x}_{m+2}, \tilde{p}_1, \dots, \tilde{p}_{m+2}).$$

For a sufficiently small positive value δ $(0 < \delta < 1)$, we construct the following distribution function

$$\bar{H}(x) = (1-\delta)H^*(x) + \delta\tilde{H}(x).$$

Denote

$$\bar{H}(x) = (\bar{x}_1, \dots, \bar{x}_{2m+3}, \bar{p}_1, \dots, \bar{p}_{2m+3}),$$

where

$$\bar{x}_i = x_i^*, \bar{p}_i = (1-\delta)p_i^*, i = 1, \dots, m+1,$$

$$\bar{x}_{m+1+i} = \tilde{x}_i, \bar{p}_{m+1+i} = \delta\tilde{p}_i, i = 1, \dots, m+2.$$

Since $H^*(x) \in K_m(X)$ and $\tilde{H}(x)$ satisfy (56), the distribution function $\bar{H}(x)$ satisfies (56) and for a sufficiently small $\delta > 0$, the following inequality holds

$$\begin{aligned} \varphi(\bar{H}) - \varphi(H^*) &= (1-\delta)\sum_{i=1}^{m+1} f_0(x_i^*)p_i^* + \delta\sum_{i=1}^{m+2} f_0(\tilde{x}_i)\tilde{p}_i - \sum_{i=1}^{m+1} f_0(x_i^*)p_i^* = \\ &= \delta\Big[\sum_{i=1}^{m+2} f_0(\tilde{x}_i)\tilde{p}_i - \sum_{i=1}^{m+1} f_0(x_i^*)p_i^*\Big] < \frac{\varepsilon_0}{8}\,. \end{aligned} \tag{58}$$

Let us construct the following distribution function

$$H'(x) = (x_1', \dots, x_{2m+3}', \bar{p}_1, \dots, \bar{p}_{2m+3}),$$

where x_i' is an arbitrary point from $U_\varepsilon(\bar{x}_i) \bigcap A_j$, if $\bar{x}_i \in A_j$, $j = 0, 1, \dots, r$; $i = 1, \dots, 2m+3$.

Since functions $f_i(x)$ are continuous on sets A_j, $i = 0, 1, \dots, m$; $j = 0, 1, \dots, r$ and $\bar{H}(x)$ satisfy (56), for sufficiently small $\varepsilon > 0$ the distribution function $H'(x)$ satisfies (56).

From the definition of sets A_j, $j = 0, 1, \ldots, r$, and the construction of the distribution function $H'(x)$, for a sufficiently small $\varepsilon > 0$, it follows that

$$\begin{aligned}\varphi(H') - \varphi(\bar{H}) &= \sum_{i=1}^{2m+3} f_0(x'_i)\bar{p}_i - \sum_{i=1}^{2m+3} f_0(\bar{x}_i)\bar{p}_i = \\ &= \sum_{i=1}^{2m+3} \bar{p}_i[f_0(x'_i) - f_0(\bar{x}_i)] < \frac{\varepsilon_0}{8}. \end{aligned} \quad (59)$$

Since $\lambda_s \to 0$ as $s \to \infty$, then for a sufficiently large s, the inequality holds $\lambda_s < \varepsilon$. Now let x'_i be a point $x^s_i \in R^s_j$ (if $\bar{x}_i \in A_j$), which is nearest to $\bar{x}_i$. Then, the distribution function $H'(x)$ satisfies (53). Since the distribution function $H_s(x)$ is a solution of the problem (51)-(53), the following inequality holds

$$\varphi(H_s) - \varphi(H') \leq 0. \quad (60)$$

From (57)-(60) for all sufficiently large s_k, we have the following inequalities

$$\begin{aligned} 0 < \varepsilon_0 &\leq \varphi(H_{s_k}) - \varphi(H^*) = \varphi(H_{s_k}) - \varphi(H') + \varphi(H') - \\ &- \varphi(\bar{H}) + \varphi(\bar{H}) - \varphi(H^*) < \frac{\varepsilon_0}{8} + \frac{\varepsilon_0}{8} = \frac{\varepsilon_0}{4}. \end{aligned} \quad (61)$$

From the contradiction obtained it follows that

$$\lim_{s\to\infty} \varphi(H_s) = \varphi(H^*). \quad (62)$$

The last limit equality (62) implies that the limit of any convergent subsequence $\{H_{s_k}\}$ belongs to the set X^*. Indeed, if the following had been the case

$$\lim_{k\to\infty} H_{s_k}(x) = H'(x) \notin X^*,$$

then the inequality $\varphi(H') > \varphi(H^*)$ would have held. Since $f_0(x)$ is lower semi-continuous, we have

$$\lim_{k\to\infty} \varphi(H_{s_k}) \geq \varphi(H') > \varphi(H^*).$$

The later contradicts to (62) and this completes the proof. $\diamond$

Case 2. Let functions $f_\nu(x)$ be upper semi-continuous, $\nu = 0, 1, \ldots, m$. In this case, the set Z defined by (48) is not closed. Therefore, infinimum (41) under constraints (42)-(43) may not be reached. However, step-functions with at most $m + 1$ steps can approximate the infinimum (41) as accurately as required. The following theorem holds.

Theorem 3.

Assume that

1. Functions $f_\nu(x)$ are upper semi-continuous and bounded on the set X, $\nu = 0, 1, 2, \ldots, m$.

2. X is a compact set from R^n.
3. $\exists \tilde{H}(x)$ such that the restrictions (55)-(56) hold.
Then

$$\lim_{s\to\infty} \varphi(H_s) = \inf_{H(x)\in K_m(X)} \varphi(H).$$

Proof. Suppose that $\varphi(H_s)$ does not converge to $\inf_{H(x)\in K_m(X)} \varphi(H)$. Then, there exists a $\varepsilon_0 > 0$ and a sequence of indexes $\{s_k\}$ such that the following inequality holds

$$\varphi(H_{s_k}) - \inf_{H(x)\in K_m(X)} \varphi(H) \geq 2\varepsilon_0 > 0\ . \tag{63}$$

Besides, there exists $H_{\varepsilon_0}(x) \in K_m(X)$, such that the following inequality holds

$$\varphi(H_{\varepsilon_0}) - \inf_{H(x)\in K_m(X)} \varphi(H) \leq \frac{\varepsilon_0}{2}. \tag{64}$$

(63) and (64) implies that

$$\varphi(H_{s_k}) - \varphi(H_{\varepsilon_0}) \geq \frac{3\varepsilon_0}{2} > 0. \tag{65}$$

For a sufficiently small positive value δ $(0 < \delta < 1)$, we construct the following distribution function

$$\tilde{H}_{\varepsilon_0}(x) = (1-\delta)H_{\varepsilon_0}(x) + \delta\tilde{H}(x).$$

Since $H_{\varepsilon_0}(x) \in K_m(X)$, and $\tilde{H}(x)$ satisfies (56), the distribution function $\tilde{H}_{\varepsilon_0}(x)$ satisfies (56) and for a sufficiently small $\delta > 0$, the following inequality holds

$$\begin{aligned} \varphi(\tilde{H}_{\varepsilon_0}) - \varphi(H_{\varepsilon_0}) &= (1-\delta)\int_X f_0(x)dH_{\varepsilon_0}(x) + \delta\int_X f_0(x)d\tilde{H}(x) - \\ &- \int_X f_0(x)dH_{\varepsilon_0}(x) = \delta\Big[\int_X f_0(x)d\tilde{H}(x) - \\ &- \int_X f_0(x)dH_{\varepsilon_0}(x)\Big] \leq \frac{\varepsilon_0}{8}. \end{aligned} \tag{66}$$

Integrals $\int_X f_\nu(x)d\tilde{H}_{\varepsilon_0}(x)$ can be approximated by sums

$$\sum_{k=1}^{N} f_\nu(x_k)p_k, \quad \sum_{k=1}^{N} p_k = 1, \quad p_k \geq 0$$

as accurate as required.

Therefore, there exists a step-function $\bar{H}(x)$ that satisfies (56) and

$$\varphi(\bar{H}) - \varphi(\tilde{H}_{\varepsilon_0}) \leq \frac{\varepsilon_0}{8}. \tag{67}$$

Since the distribution function $\bar{H}(x)$ is a step-function, the point $Q = (\varphi(\bar{H}), \psi_1(\bar{H}), \ldots, \psi_m(\bar{H})) \in coZ$ and can be represented as a convex combination of at most $m+2$ points from Z. Therefore, we can suggest that $\bar{H}(x)$ is a step-function with at most $m+2$ steps.

Denote

$$\bar{H}(x) = (\bar{x}_1, \dots, \bar{x}_{m+2}, \bar{p}_1, \dots, \bar{p}_{m+2})$$

and

$$H'(x) = (x'_1, \dots, x'_{m+2}, \bar{p}_1, \dots, \bar{p}_{m+2}),$$

where x'_i is an arbitrary point from the set $U_\varepsilon(\bar{x}_i) \bigcap X$, $i = 1, \dots, m+2$. Since the distribution function $\bar{H}(x)$ satisfies (56) and all functions $f_i(x)$, $i = 1, \dots, m$, are upper semi-continuous at points $\bar{x}_i$, for a sufficiently small $\varepsilon > 0$, the following inequality holds

$$\begin{aligned} \psi_i(H') &= \sum_{j=1}^{m+2} f_i(x'_j)\bar{p}_j = \sum_{j=1}^{m+2} f_i(x'_j)\bar{p}_j - \sum_{j=1}^{m+2} f_i(\bar{x}_j)\bar{p}_j + \psi_i(\bar{H}) = \\ &= \sum_{j=1}^{m+2} \bar{p}_j[f_i(x'_j) - f_i(\bar{x}_j)] + \psi_i(\bar{H}) < a_i, \quad i = 1, \dots, m. \end{aligned} \tag{68}$$

Therefore, $H'(x) \in K_m(X)$, and since the function $f_0(x)$ is upper semi-continuous at the points $\bar{x}_i$, $i = 1, \dots, m+2$, the following inequality holds

$$\begin{aligned} \varphi(H') - \varphi(\bar{H}) &= \sum_{j=1}^{m+2} f_0(x'_j)\bar{p}_j - \sum_{j=1}^{m+2} f_0(\bar{x}_j)\bar{p}_j = \\ &= \sum_{j=1}^{m+2} \bar{p}_i[f_0(x'_j) - f_0(\bar{x}_j)] < \frac{\varepsilon_0}{8}. \end{aligned} \tag{69}$$

Since $\lambda_s \to 0$ as $s \to \infty$, for a sufficiently large s the inequality holds $\lambda_s < \varepsilon$. Now let x'_i be a point $x_i^s \in R^s$, which is nearest to $\bar{x}_i$. Then, the distribution function $H'(x)$ satisfies (53).

Since the distribution function $H_s(x)$ is a solution of the problem (51)-(53), then the following inequality holds

$$\varphi(H_s) - \varphi(H') \leq 0. \tag{70}$$

From (65)-(70) for all sufficiently large s_k, we have the following inequalities

$$\begin{aligned} 0 < \frac{3\varepsilon_0}{2} &\leq \varphi(H_{s_k}) - \varphi(H_{\varepsilon_0}) = \varphi(H_{s_k}) - \varphi(H') + \varphi(H') - \varphi(\bar{H}) + \\ &+ \varphi(\bar{H}) - \varphi(\tilde{H}_{\varepsilon_0}) + \varphi(\tilde{H}_{\varepsilon_0}) - \varphi(H_{\varepsilon_0}) < \\ &< \frac{\varepsilon_0}{8} + \frac{\varepsilon_0}{8} + \frac{\varepsilon_0}{8} = \frac{3\varepsilon_0}{8}. \end{aligned} \tag{71}$$

The contradiction obtained implies that

$$\lim_{s \to \infty} \varphi(H_s) = \inf_{H(x) \in K_m(X)} \varphi(H).$$

The theorem is proved. $\diamond$

5 Optimization of linear functional under fixed moments

Theorems in the previous section provides the algorithm only when the set $K_m(X)$ contains an inner point $\tilde{H}(x)$. However, it is often required to optimize (41) under the condition that m first moments of a distribution function $H(x)$ are fixed. This is the case when the third condition of Theorems 2 and 3 are violated. Since this class of problems is very important, let us investigate in detail the numerical technique for solving such problems.

Let us consider the following problem: it is required to minimize

$$\varphi(H) = \int_a^b f_0(x)dH(x) \tag{72}$$

subject to

$$\int_a^b dH(x) = 1,$$

$$\int_a^b x^i dH(x) = \mu_i, \tag{73}$$

$$i = 1, \dots, m, \quad -\infty < a < b < \infty.$$

Denote by $K_m(a,b)$ the set of distribution functions that satisfy (73). We consider only the case when $f_0(x)$ is lower semi-continuous. As in the previous section, it can be shown that a solution of the problem (72)-(72) is achieved at a step-function with at most $m+1$ steps.

Algorithm.

On the sth iteration, construct partition $R_s = (x_1, x_2, \dots, x_{n_s})$ of the segment $[a,b]$ such that intervals of length r_s with centers in points $x_i \in [a,b]. i = 1, \dots, n_s$ would cover the segment $[a,b]$; $a = x_1 \le x_2 \le \dots \le x_{n_s} = b$; $R_s \subset R_{s+1}$ and $r_s \to 0$ as $s \to \infty$.

Then, solve the following linear programming problem:

$$\sum_{j=1}^{n_s} f_0(x_j)p_j \to \min \tag{74}$$

subject to

$$\sum_{j=1}^{n_s} p_j = 1,$$

$$\sum_{j=1}^{n_s} x_j^i p_j = \mu_i \quad i = 1, 2, \dots, m, \tag{75}$$

$$p_j \ge 0, \qquad x_j \in R_s, \qquad j = 1, 2, \dots, n_s.$$

Let $p^s = (p_1^s, p_2^s, \ldots, p_{n_s}^s)$ be a solution of the problem (74)-(75). Since the number of basic variables in the problem (74)-(75) equals $m+1$, at most $m+1$ components of the vector p^s can be nonzero. Without any loss of generality, we can assume that the first $m+1$ components $p_1^s, p_2^s, \ldots, p_{m+1}^s$ are nonzero. Therefore, we can present the solution of the problem (74)-(75) as follows

$$H_s(x) = (x_1^s, \ldots, x_{m+1}^s, p_1^s, \ldots, p_{m+1}^s), \tag{76}$$

where the basic variable p_i^s corresponds to point x_i^s and appropriate basic column

$$(1, x_i^s, (x_i^s)^2, \ldots, (x_i^s)^m)'.$$

On the other hand, (76) can be considered as a presentation of the distribution function, where the step-function $H_s(x)$ has the step p_i^s at the point x_i^s, $i = 1, 2, \ldots, m$.

The solution $H_s(x)$ found at the s-th iteration will be used as initial basic variables in solving the problem (74)- (75) at the $s+1$-th iteration.

Before formulating the algorithm discussed above, let us consider some properties of the moment space (see [7], [8], [6]).

Let D denote the collection of all distribution functions whose support is the segment $[a, b]$, and D_A denote the collection of all step-functions whose support is the segment $[a, b]$. Let $I(t - t_0)$ denote the following one-step distribution function

$$I(t - t_0) = \begin{cases} 1, \text{ if } & t \geq t_0 \\ 0, \text{ if } & t < t_0. \end{cases}$$

One-step distribution functions are extreme points of D_A. Any step-function can be presented as a convex combination of one-step distribution functions.

Definition 1. Moment space D^m is said to be the set of points $x = (x_1, \ldots, x_m) \in E_m$ whose coordinates are moments $\mu_1(H), \mu_2(H), \ldots, \mu_m(H)$ for at least one distribution function from D.

A point in D^m, which corresponds to the one-step distribution function $I(t - t_1)$ is denoted by $x(t_1) = (t_1, t_1^2, \ldots, t_1^m)$.

Let C^m denote a curve $x(t_1)$, $a \leq t_1 \leq b$. The following well known theorems hold.

Theorem 4.

D^m is a convex set.

Theorem 5.

A set of extreme points of D^m for $m \geq 2$ coincides with C^m.

Theorem 7.

Presentation of a point $x \in D^m$ in the form of a convex combination of extreme points is unique if and only if x belongs to boundary of D^m.

From Theorem 7 the following theorem follows.

Theorem 8.

The extreme point of D^m and only it corresponds to the unique distribution function from D.

It is obvious that for any $k \geq 1$, any inner point $x \in D^m$ can be presented as a convex combination of $r\,(r > k)$ points from C^m. Hence, for any k and any system of moments $(\mu_1, \ldots, \mu_m)$, which is an inner point of D^m, for some $r > k$ there exists step-function $H(x)$ with r steps and moments $\mu_1, \ldots, \mu_m$.

Theorem 8 implies that if $(\mu_1, \ldots, \mu_m)$ is a boundary point of D^m, then, the problem (72)-(73) is reduced to the searching of a unique distribution function that satisfies (73). We assume that $(\mu_1, \ldots, \mu_m)$ is an inner point of D^m. The following theorem holds.

Theorem 9.

Assume that

1. Function $f_0(x)$ is lower semi-continuous and bounded on the segment [a,b] with a finite number of points of discontinuity.

2. Point $(\mu_1, \ldots, \mu_m)$ is an inner point of the moment space D^m.

Then,

$$\lim_{s\to\infty} \varphi(H_s) = \varphi(H^*).$$

and limit of any convergent subsequence $\{H_{s_k}\}$ of the sequence $\{H_s\}$ belongs to the set X^*, where

$$X^* = \{H^*(x) : \varphi(H^*) = \min_{H(x)\in K_m(a,b)} \varphi(H)\}$$

Proof. Suppose that $\varphi(H_s)$ does not converge to $\varphi(H^*)$. Then, there exists a $\varepsilon_0 > 0$ and a sequence of indexes $\{s_k\}$ such that the following inequality holds

$$\varphi(H_{s_k}) - \varphi(H^*) \geq \varepsilon_0 > 0. \tag{77}$$

The distribution function $H^*(x) \in X^*$ has k steps $(k \leq m+1)$. As it was pointed out above, there exists the distribution function $\tilde{H}(x)$, which has moments $(\mu_1, \ldots, \mu_m)$ and r steps $(r + k \geq m + 1)$.

Without a loss of generality, we can assume that points in which distribution functions $H^*(x)$ and $\tilde{H}(x)$ have steps are different. Consider the following distribution function

$$\bar{H}(x) = (1-\alpha)H^*(x) + \alpha\tilde{H}(x),$$

where α is a sufficiently small positive value $(0 < \alpha < 1)$.

Denote

$$H^*(x) = (x_1^*, \ldots, x_k^*, p_1^*, \ldots, p_k^*),$$

$$\tilde{H}(x) = (\tilde{x}_1, \ldots, \tilde{x}_r, \tilde{p}_1, \ldots, \tilde{p}_r),$$

$$\bar{H}(x) = (\bar{x}_1, \ldots, \bar{x}_{k+r}, \bar{p}_1, \ldots, \bar{p}_{k+r}),$$

where

$$\bar{x}_i = x_i^*, \quad \bar{p}_i = (1-\alpha)p_i^*,\ p_i^* > 0,\ i = 1, \ldots, k,$$

$$\bar{x}_{k+i} = \tilde{x}_i, \quad \bar{p}_{k+i} = \alpha \tilde{p}_i,\ \tilde{p}_i > 0,\ i = 1, \ldots, r.$$

Since the function $f_0(x)$ is bounded on $[a, b]$, for a sufficiently small $\alpha > 0$, the following inequality holds

$$\begin{aligned} \varphi(\bar{H}) - \varphi(H^*) &= (1-\alpha)\sum_{i=1}^{k} f_0(x_i^*)p_i^* + \alpha \sum_{i=1}^{r} f_0(\tilde{x}_i)\tilde{p}_i - \sum_{i=1}^{k} f_0(x_i^*)p_i^* = \\ &= \alpha\Big[\sum_{i=1}^{r} f_0(\tilde{x}_i)\tilde{p}_i - \sum_{i=1}^{k} f_0(x_i^*)p_i^*\Big] < \frac{\varepsilon_0}{8}\,. \end{aligned} \tag{78}$$

Since distribution functions $H^*(x)$ and $\tilde{H}(x)$ have the same m first moments, the distribution function $\bar{H}(x)$ satisfies (73)

$$\sum_{j=1}^{m+1} \bar{p}_j = 1 - \sum_{j=m+2}^{k+r} \bar{p}_j,$$

$$\sum_{j=1}^{m+1} \bar{x}_j^i \bar{p}_j = \mu_i - \sum_{j=m+2}^{k+r} \bar{x}_j^i \bar{p}_j, \tag{79}$$

$$i = 1, \ldots, m, \quad \bar{p}_j > 0, \quad j = 1, \ldots, k + r.$$

Since $det B \neq 0$, where

$$B = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \bar{x}_1 & \bar{x}_2 & \ldots & \bar{x}_{m+1} \\ \ldots & \ldots & \ldots & \ldots \\ \bar{x}_1^m & \bar{x}_2^m & \ldots & \bar{x}_{m+1}^m, \end{pmatrix}$$

from equations (79), we can express $p_1, p_2, \ldots, p_{m+1}$ as functions of $(x_1, \ldots, x_{k+r}, p_{m+2}, \ldots, p_{k+r})$, which are continuous in the point $(\bar{x}_1, \ldots, \bar{x}_{k+r}, \bar{p}_{m+2}, \ldots, \bar{p}_{k+r})$.

Since

$$\bar{p}_i = g_i(\bar{x}_1, \ldots, \bar{x}_{k+r}, \bar{p}_{m+2}, \ldots, \bar{p}_{k+r}) > 0,$$

$$i = 1, \ldots, m+1,$$

then for sufficiently small $\delta > 0$ and any $x_i' \in U_\delta(\bar{x}_i) \bigcap [a, b]$

$$p_i' = g_i(x_1', \ldots, x_{k+r}', \bar{p}_{m+2}, \ldots, \bar{p}_{k+r}) > 0,$$

$$i = 1, \ldots, m+1,$$

and the distribution function equals

$$H'(x) = (x_1', \ldots, x_{k+r}', p_1', \ldots, p_{k+r}'),$$

where

$$p_i' = g_i(x_1', \dots, x_{k+r}', \bar{p}_{m+2}, \dots, \bar{p}_{k+r}),$$

$$i = 1, \dots, m+1,$$

$$p_i' = \bar{p}_i, \; i = m+2, \dots, k+r,$$

satisfies (73).

Let us select points x_i' as follows:

if $f_0(\bar{x}_i) = f_0(\bar{x}_i + 0)$, then x_i' - any point from $(\bar{x}_i, \bar{x}_i + \delta) \bigcap [a, b]$;

if $f_0(\bar{x}_i) = f_0(\bar{x}_i - 0)$, then x_i' - any point from $(\bar{x}_i - \delta, \bar{x}_i) \bigcap [a, b]$.

Then, the first condition of the theorem for sufficiently small $\delta > 0$ implies

$$|f_0(x_i') - f_0(\bar{x}_i)| < \frac{\varepsilon_0}{16C(k+r)}, \tag{80}$$

$$i = 1, \dots, k+r.$$

From the continuity of functions $g_i(x_1, \dots, x_{k+r}, \bar{p}_{m+2}, \dots, \bar{p}_{k+r})$ in the point $(\bar{x}_1, \dots, \bar{x}_{k+r}, \bar{p}_{m+2}, \dots, \bar{p}_{k+r})$ and from (80), for a sufficiently small $\delta > 0$, we have

$$\begin{aligned}
\varphi(H') \;&-\; \varphi(\bar{H}) = \sum_{i=1}^{m+1} f_0(x_i') g_i(x_1', \dots, x_{k+r}', \bar{p}_{m+2}, \dots, \bar{p}_{k+r}) + \\
&+ \sum_{i=m+2}^{k+r} f_0(x_i')\bar{p}_i - \sum_{i=1}^{m+1} f_0(\bar{x}_i) g_i(\bar{x}_1, \dots, \bar{x}_{k+r}, \bar{p}_{m+2}, \dots, \bar{p}_{k+r}) - \\
&- \sum_{i=m+2}^{k+r} f_0(\bar{x}_i)\bar{p}_i = \sum_{i=1}^{m+1} [f_0(x_i') - f_0(\bar{x}_i)] g_i(x_1', \dots, x_{k+r}', \bar{p}_{m+2}, \dots, \bar{p}_{k+r}) + \\
&+ \sum_{i=m+2}^{k+r} [f_0(x_i') - f_0(\bar{x}_i)]\bar{p}_i + \sum_{i=1}^{m+1} f_0(\bar{x}_i)[g_i(x_1', \dots, x_{k+r}', \bar{p}_{m+2}, \dots, \bar{p}_{k+r}) - \\
&- g_i(\bar{x}_1, \dots, \bar{x}_{k+r}, \bar{p}_{m+2}, \dots, \bar{p}_{k+r})] < \\
&< C(k+r)\frac{\varepsilon_0}{16C(k+r)} + \frac{\varepsilon_0}{16} = \frac{\varepsilon_0}{8}.
\end{aligned} \tag{81}$$

Since $r_s \to 0$ as $s \to \infty$, for a sufficiently large s, the inequality $r_s < \delta$ holds. Select point x_i' by using the following rule.

If the function $f_0(x)$ is left semi-continuous in the point $\bar{x}_i$, then select point $x_i^s \in R_s$ nearest from the left to $\bar{x}_i$

If the function $f_0(x)$ is right semi-continuous in the point $\bar{x}_i$, then select point $x_i^s \in R_s$ nearest from the right to $\bar{x}_i$.

Thus, the distribution function $H'(x)$ satisfies (75). Since the distribution function $H_s(x)$ is a solution of the problem (74)-(75), the following inequality holds

$$\varphi(H_s) - \varphi(H') \le 0. \tag{82}$$

From (77)-(82) for all sufficiently large s_k, we have

$$\begin{aligned} 0<\varepsilon_0 &\leq \varphi(H_{s_k})-\varphi(H^*)=\varphi(H_{s_k})-\varphi(H')+\varphi(H')- \\ &- \varphi(\bar{H})+\varphi(\bar{H})-\varphi(H^*)<\frac{\varepsilon_0}{8}+\frac{\varepsilon_0}{8}=\frac{\varepsilon_0}{4}. \end{aligned} \tag{83}$$

From the obtained contradiction it follows that

$$\lim_{s\to\infty}\varphi(H_s)=\varphi(H^*). \tag{84}$$

(84) implies that the limit of any convergent subsequence $\{H_{s_k}\}$ belongs to the set X^* and this completes the proof. $\diamond$

6 Optimization of linear-fractional functional under constraints of inequality type

Now, let us consider the following problem

$$\varphi(H)=\frac{\int_X g_1(x)dH(x)}{\int_X g_2(x)dH(x)}\to \inf_{H(X)\in K_m(X)} \tag{85}$$

where $K_m(X)$ is the set of distribution functions which satisfy (42)-(43). Let us denote $J_1(H)=\int_X g_1(x)dH(x)$ and $J_2(H)=\int_X g_2(x)dH(x)$. We assume that $J_2(H)>\gamma>0$ for all distribution functions $H(x)\in K_m(X)$, where γ is a fixed positive number. The following theorem holds.

Theorem 10. The problem (85) is equivalent to a problem of searching for a number t_* such that the following equality holds

$$\inf_{H\in K_m(X)}[J_1(H)-t_*J_2(H)]=0. \tag{86}$$

Proof. Let (86) holds. Then, for any distribution function $H(x)\in K_m(X)$, we have

$$J_1(H)-t_*J_2(H)\geq 0.$$

Therefore, for $\forall\, H(x)\in K_m(X)$ we have

$$t_*\leq\frac{J_1(H)}{J_2(H)}. \tag{87}$$

Besides, for $\forall\,\varepsilon>0$ there exists distribution function $H_\varepsilon(x)\in K_m(X)$ such that

$$J_1(H_\varepsilon)-t_*J_2(H_\varepsilon)\leq\varepsilon\gamma.$$

Therefore,

$$t_*\geq\frac{J_1(H_\varepsilon)}{J_2(H_\varepsilon)}-\varepsilon\frac{\gamma}{J_2(H_\varepsilon)}\geq\frac{J_1(H_\varepsilon)}{J_2(H_\varepsilon)}-\varepsilon. \tag{88}$$

From (87)-(88), we have

$$t_* = \inf_{H(x)\in K_m(X)} \frac{J_1(H)}{J_2(H)}.$$

The same reasoning can be used to prove the theorem in the reverse direction.

From the results of section 1, it follows that if functions $g_1(x), f_1(x), \dots, f_m(x)$ are lower semi-continuous and the function $g_2(x)$ is upper semi-continuous, then, the minimum of the functional (86) is achieved on a step-function with at most $m+1$ steps. Since the problem (86) is equivalent to the problem (85), the minimum of the functional (85) is achieved at a step-function with at most $m+1$ steps.

Suppose that functions $g_1(x), g_2(x), f_i(x), i = 1, \dots, m$, have a finite number of surfaces of discontinuity $A_1, A_2, \dots, A_q$. Suppose that these functions are continuous on $A_1, A_2, \dots, A_q$.

Algorithm.

On the sth iteration, construct partitions $R_s^i = (x_{1,i}^s, x_{2,i}^s, \dots, x_{l_s,i}^s)$ of sets $A_i, i = 0, 1, \dots, q$, such that $A_i \subset \bigcup_{j=1}^{l_s} U_{r_s}(x_{j,i}^s)$, $R_s^i \subset R_{s+1}^i$, $i = 0, 1, \dots, q$; $r_s \to 0$ as $s \to \infty$, where $A_0 = X \setminus \bigcup_{i=1}^q A_i)$; $U_{r_s}(x_{j,i}^s) = \{x : \|x - x_{j,i}^s\| < r_s\}$.

Let $R_s = \bigcup_{i=0}^q R_s^i$, $R_s = (x_1^s, x_2^s, \dots, x_{n_s}^s)$, where $n_s = (q+1) \cdot l_s$, $x_{(j+1)\cdot l_s+i}^s = x_{i,j}^s$, $i = 1, 2, \dots, l_s$; $j = 0, 1, \dots, q$.

Then, let us solve the following linear programming problem

$$\sum_{j=1}^{n_s} [g_1(x_j) - t_{s-1} g_2(x_j)] p_j \to \min \tag{89}$$

subject to

$$\sum_{j=1}^{n_s} p_j = 1, \tag{90}$$

$$\sum_{j=1}^{n_s} f_i(x_j) p_j \le a_i \qquad i = 1, 2, \dots, m, \tag{91}$$

$$p_j \ge 0, \qquad x_j \in R_s, \qquad j = 1, 2, \dots, n_s.$$

Let a distribution function $H_s(x)$ be a solution of this problem. Then, let us calculate

$$t_s = \frac{J_1(H_s)}{J_2(H_s)} \tag{92}$$

and go to the $s+1$th iteration.

As an initial approximation, take the value

$$t_0 = \frac{J_1(H_0)}{J_2(H_0)},$$

where H_0 is an arbitrary distribution function from $K_m(X)$ with at most $m+1$ steps.

The following theorem holds.

Theorem 11.

Assume that

1. Functions $g_1(x), f_i(x)$, $i = 1, \dots, m$, are lower semi-continuous and $g_2(x)$ is upper semi-continuous and bounded on the set X.

2. X is a compact set from R^n.

3. $\exists \tilde{H}(x)$ such that the following restrictions hold

$$\int_X d\tilde{H}(x) = 1,$$

$$\psi_i(\tilde{H}) = \int_X f_i(x) d\tilde{H}(x) < a_i, \qquad (93)$$

$$i = 1, 2, \dots, m,$$

Then, limit of any convergent subsequence $\{H_{s_k}\}$ of the sequence $\{H_s\}$ belongs to the set X^*, where

$$X^* = \{H^*(x) : \varphi(H^*) = \min_{H(x) \in K_m(X)} \varphi(H)\}$$

and

$$\lim_{s \to \infty} t_s = \min_{H(x) \in K_m(X)} \frac{\int_X g_1(x) dH(x)}{\int_X g_2(x) dH(x)}$$

Proof.

The sequence $t_1, \dots .t_s, \dots$ is lower bounded since for all s the following inequality holds

$$t_s \geq \inf_{H(x) \in K_m(X)} \frac{J_1(H}{J_2(H)}.$$

Besides, the sequence is non-increasing. Indeed, from (89) and (92), we have

$$J_1(H_s) - t_{s-1} J_2(H_s) \leq J_1(H_{s-1}) - t_{s-1} J_2(H_{s-1}) = 0.$$

Since, according to our assumption, $J_2(H_s) \geq \gamma > 0$,

$$t_{s-1} \geq \frac{J_1(H_s)}{J_2(H_s)} = t_s.$$

Therefore, there exists the limit

$$\lim_{s \to \infty} t_s = t_*.$$

If we show that

$$\min_{H(x) \in K_m(X)} [J_1(H) - t_* J_2(H)] = J_1(H^*) - t_* J_2(H^*) = 0\,,$$

then

$$t_* = \min_{H(x) \in K_m(X)} \frac{J_1(H)}{J_2(H)} = \frac{J_1(H^*)}{J_2(H^*)}.$$

Suppose the opposite is true, then

$$J_1(H^*) - t_* J_2(H^*) < 0.$$

Consequently, there exists a value $\varepsilon_0 > 0$ such that

$$J_1(H^*) - t_* J_2(H^*) \leq -\varepsilon_0. \tag{94}$$

According to our assumptions and the structure of the algorithm, we can show in a way similar to that of section 1 that for sufficiently large s there exist distribution functions $\bar{H}_s(x) \in K_m(R_s)$, such that

$$J_1(H^*) - t_* J_2(H^*) > J_1(\bar{H}_s) - t_* J_2(\bar{H}_s) - \frac{\varepsilon_0}{8}, \tag{95}$$

where $K_m(R_s) \subset K_m(X)$ and $K_m(R_s)$ - the set of all distribution functions whose support belong to R_s.

Since $t_s \to t_*$ as $s \to \infty$, for sufficiently large s, we have

$$J_1(\bar{H}_s) - t_* J_2(\bar{H}_s) > J_1(\bar{H}_s) - t_{s-1} J_2(\bar{H}_s) - \frac{\varepsilon_0}{8}. \tag{96}$$

Since $H_s(x)$ is a solution of the problem

$$\min_{H(x) \in K_m(R_s)} [J_1(H) - t_{s-1} J_2(H)],$$

for sufficiently large numbers s, the following inequality holds

$$\begin{aligned} J_1(\bar{H}_s) &- t_{s-1} J_2(\bar{H}_s) \geq J_1(H_s) - t_{s-1} J_2(H_s) \geq \\ &\geq J_1(H_s) - t_s J_2(H_s) - \frac{\varepsilon_0}{8} = -\frac{\varepsilon_0}{8} \end{aligned} \tag{97}$$

Therefore, we have

$$\begin{aligned} 0 > -\varepsilon_0 &\geq J_1(H^*) - t_* J_2(H^*) > J_1(\bar{H}_s) - t_* J_2(\bar{H}_s) - \frac{\varepsilon_0}{8} > \\ &> J_1(\bar{H}_s) - t_{s-1} J_2(\bar{H}_s) - \frac{2\varepsilon_0}{8} > J_1(H_s) - t_{s-1} J_2(H_s) - \frac{2\varepsilon_0}{8} > -\frac{3\varepsilon_0}{8} \end{aligned} \tag{98}$$

From the obtained contradiction, we have

$$\min_{H(x) \in K_m(X)} [J_1(H) - t_* J_2(H)] = 0.$$

Therefore,

$$t_* = \min_{H(x) \in K_m(X)} \frac{J_1(H)}{J_2(H)}.$$

Thus,

$$\lim_{s \to \infty} t_s = \lim_{s \to \infty} \varphi(H_s) = \min_{H \in K_m(X)} \varphi(H). \tag{99}$$

(99) implies that the limit of any convergent subsequence $\{H_{s_k}\}$ of the sequence $\{H_s\}$ belongs to the set X^*. The theorem is proved. $\diamond$

Remarks.

1. If at least one of the functions $f_i(x)\, i = 1, \dots, m$, or the function $\frac{g_1(x)}{g_2(x)}$ is upper semi-continuous, then the algorithm converges to the infinimum of the functional $\varphi(H)$.

2. The algorithm can be slightly modified to optimize the functional (85) under m fixed first moments of distribution functions $H(x)$. For convergence of appropriate algorithm, it is necessary that the point $(\mu_1, \dots, \mu_m)$ be an inner point of the moment space D^m. If the function $\frac{g_1(x)}{g_2(x)}$ is lower semi-continuous, then the algorithm converges to the minimum of the functional (85), and if it is upper semi-continuous, then the algorithm converges to infinimum of the (85).

Proof of these statements can be carried out in the same manner as in the previous sections.

7 Results of numerical calculations

Further, for a particular application, we calculate upper and lower bounds for Bayes estimates corresponding to the same available prior information. The case study was done for binomial sampling. We suppose that available prior information allows us to estimate the values of the first and the second moments of the prior distribution.

In order to calculate the upper and lower bounds for the Bayes estimates, we should solve the problems (26), (22) - (23) and (27), (22) - (23), where $a = 0, \quad b = 1, \quad i = 2$ and μ_1 is the first moment and μ_2 is the second moment of the prior distribution. Given μ_1 and σ^2, the second moment is calculated as follows

$$\mu_2 = mu_1^2 + \sigma^2.$$

Calculations were performed for different values of the mean μ_1 and variance σ^2 of the prior distribution, the number of trials n, and the number of failures s. Upper and lower bounds for Bayes estimates calculated for fixed values of these parameters were compared with the Bayes estimate (15) derived under the beta prior distribution, which fits selected values of parameters.

Range of values of the parameters was chosen in such a manner that they were close to that encountered in PSA of Nuclear Power Plants.

Table 3 gives results of numerical calculations of upper and lower bounds for Bayes estimates derived under any prior distribution having the same mean $\mu = 0.01$ and variance $\sigma^2 = 1.0E - 4$. The number of trials was set $n = 220$ and the number of failures s is variated from 0 to 10. Obtained results are compared with the Bayes estimate (beta estimate) derived under the beta prior distribution, which has the same mean and variance.

Table 3: Upper and lower bounds for Bayes estimates and Bayes estimate derived under beta prior distribution (beta estimate) for $\mu = 0.01, \quad \sigma^2 = 1.0E-4$ and $n = 220$

s	*lower bound*	*beta estimate*	*upper bound*
0	4.23E-04	3.08E-03	9.90E-03
1	3.46E-03	6.23E-03	1.28E-02
2	5.85E-03	9.37E-03	1.73E-02
3	7.20E-03	1.25E-02	2.09E-02
4	7.97E-03	1.57E-02	2.33E-02
5	8.42E-03	1.88E-02	2.73E-02
6	8.79E-03	2.19E-02	3.29E-02
7	8.97E-03	2.51E-02	4.08E-02
8	9.06E-03	2.82E-02	5.09E-02
9	9.30E-03	3.14E-02	6.24E-02
10	9.47E-03	3.45E-02	7.56E-02

Table 3 demonstrates that for small numbers of failures ($s = 0, 1$), the range between upper and lower bounds is large and the discrepancy between upper bounds and Bayes estimates derived under beta prior distribution is the largest.

The dependence of upper and lower bounds as well as the beta estimate of the variance is presented in Table 4.

Table 4: Dependence of upper and lower bounds as well as the beta estimate of the variance for $\mu = 0.01$,$n = 100$ and $s = 1$

Variance	1.00E-06	5.00E-06	1.00E-05	5.00E-05	1.00E-04
Lower bound	9.9715E-03	9.8565E-03	9.7104E-03	8.5987E-03	7.3750E-03
Beta estimate	1.0000E-02	1.0000E-02	1.0000E-02	1.0000E-02	1.0000E-02
Upper bound	1.0076E-02	1.0393E-02	1.0805E-02	1.3428E-02	1.4983E-02

Table 4 demonstrates that the range between upper and lower bounds as well as the discrepancy between upper bounds and Bayes estimates derived under the beta prior distribution increases with the increase of variance σ^2.

Results of calculations presented in the tables show that there are ranges of parameters of practical significance where Bayes estimates are sensitive to the type of a prior distribution function that has the same first two moments. In these cases, using the beta distribution as a prior may lead to a significant underestimation of the probability of failure in binomial sampling. For this reason, the conservative Robust Bayes estimate should be used.

References

[1] Yu.Ermoliev, Method for Stochastic Programming in Randomized Strategies, Kibernetica,1 (1970), pp.3-9. (In Russian).

[2] Yu.Ermoliev,Methods of Stochastic Programming, Nauka, Moscow, 1976. (In Russian)

[3] A.N. Golodnikov and L.S. Stoikova, The Determination of the Optimal Period of Preventive Replacement on the Basis of Information on Mathematical Expectation and Time Dispersion of the Trouble-Free Operation of a System. (Russian) Kibernetika (Kiev) 1978, No.3, 110-118, English translation: Cybernetics 14 (1978), No.3, 431-440 (1979).

[4] A.N. Golodnikov, Minimax Approach to Bayes Estimation. Operation Research (Models, Systems, Solutions) No 7, pp. 36-41, 1, Akad. Nauk. SSSR, Vycisl. Centr. Moscow (1979) (In Russian).

[5] D.R. Insua and F. Ruggeri, Eds., Robust Bayesian Analysis. Lecture Notes in Statistics. Springer Verlag (2000).

[6] S.Karlin and W.J.Studden, Tchebycheff Systems: With Applications in Analysis and Statistics. Wiley Interscience. New York, 1966.

[7] J.H.B.Kemperman, The General Moment Problem, a Geometric Approach. Ann. Math. Statist., 39 (1968), pp.93-122.

[8] M.Krein and A.Nudelman, The Markov Moment Problem and Extremal Problems, Trans. Math. Monographs, American Mathematical Society, Providence, RI,1977

[9] Harry F. Martz and Ray A. Waller. Bayesian reliability Analysis. Krieger Publishing Company, Malabar, Florida, 1991.

[10] M. Lavine, at al., Linearization techniques in Bayesian Robustness. In.: Robust Bayesian Analysis. D.R. Insua and F. Ruggeri, Eds., Lecture Notes in Statistics. Springer Verlag (2000).

[11] R. J-B. Wets, Statistical estimation from an optimization viewpoint. Annals of Operations Research 85 (1999)79-101.