Applications of Mathematics

Pavel Charamza

Recursive estimates of quantile based on 0-1 observations

Applications of Mathematics, Vol. 37 (1992), No. 3, 173–192

Persistent URL: http://dml.cz/dmlcz/104502

37 (1992) APPLICATIONS OF MATHEMATICS No. 3, 173–192

RECURSIVE ESTIMATES OF QUANTILE BASED ON 0–1 OBSERVATIONS

PAVEL CHARAMZA

(Received November 20, 1990)

Summary. The objective of this paper is to introduce some recursive methods that can be used for estimating an *LD-50* value. These methods can be used more generally for the estimation of the γ-quantile of an unknown distribution provided we have 0–1 observations at our disposal. Standard methods based on the Robbins-Monro procedure are introduced together with different approaches of Wu or Mukerjee. Several examples are also mentioned in order to demonstrate the usefulness of the methods presented.

Keywords: *LD-50* value, quantile, stochastic approximation, isotonic regression

PROBLEM FORMULATION

The problem is formulated as follows: on the level x ($x \in \mathbf{R}$) a random variable Y_x is observed, distributed according to the 0-1 law with an unknown parameter $F(x)$. Having observations on different levels x we want to estimate the solution of the equation $F(x) = \gamma$, where $\gamma \in (0, 1)$ is a given number. The function F is usually a distribution function; however, the fulfillment of conditions

$$F(x) < \gamma, \quad x < x^*,$$
$$F(x) > \gamma, \quad x > x^*$$

for some real number x^* is sufficient for a majority of algorithms.

Examples

First we give some practical examples that can be mathematically formulated as the above stated problem.

Example 1. *Reliability of material*

Let X be a random variable indicating the power under which some material is defected. Performing an experiment on a power level x we get the result Y_x with the following distribution:

$$Y_x = 0 \quad \text{(no defect) with probability} \quad 1 - F(x),$$
$$Y_x = 1 \quad \text{(defect) with probability} \quad F(x),$$

where F is the distribution function of the variable X.

Our task is to determine the maximum power that does not effect the structure of a high percentage of material, i.e. to find the solution of the equation $F(x) = \gamma$, where $\gamma \ll 1$.

Example 2. *Explosive testing*

The random variable X indicates the pressure under which an explosive will explode. The result on the level x is explosion or no explosion with probability $F(x)$ or $1 - F(x)$, respectively, where F is the distribution function of X as in Example 1. The task is to find the pressure under which $100\gamma\,\%$ of the explosive will explode. In this case $\gamma \gg 0$ is usually taken.

Example 3. *Biological experiments*

We apply a dose x of a substance to an experimental animal. The result of the experiment is 1 if the animal somehow responds to this dose. Otherwise the result is 0. The response of the individual appears with probability $F(x)$, taking F as the distribution function of the random variable that expresses the amount of the dose necessary for the response. We want to estimate the level of the dose on which $100\gamma\,\%$ of individuals response. The value for γ is usually taken as 0.5 and the corresponding level is then called *LD-50*.

Example 4. *Range of feromon*

Let $G(x)$ denote the probability of the event that insects will reach the feromon being x units far from it. The marked insects either reach the investigated feromon or not. We look for the γ-quantile of the distribution of the random variable X denoting the maximum distance from which the insects reach the feromon. The distribution of this variable is clearly given by the distribution function $F \stackrel{\text{def}}{=} 1 - G$. Thus we look for the solution of the equation $F(x) = \gamma$.

Example 5. *Psychological testing*

The problem is to construct a sequence of easier and more complicated questions in such a way that the complexity of these questions gradually increases and respects the individual abilities of the person investigated. After some time the individual abilities are judged using the level of complexity of the questions. The following model is used in psychology:

The examiner has a set of questions. The answers can be classified as "good" or "bad". We denote the answer to the question i by the symbol u_i; $u_i = 1$ if the answer is correct, $u_i = 0$ otherwise. The question is characterized by the so called characteristic curve $P_i(\theta) = P[u_i = 1|\theta]$, where θ is a real parameter denoting the individual abilities. Functions $P(\theta)$ are assumed to be continuous and increasing. The logistic curves are usually considered. The complexity of the question is defined as the "γ-quantile" of the function $P_i(\theta)$, more precisely as the solution of the equation $P_i(x) = \gamma$, where $\gamma \in (0,1)$. We put questions to the individual on different levels i_x. Taking into account his previous answers we try to find the complexity x^* such that the questions on this complexity level are answered correctly with probability γ. Psychologists consider this value to be the measure of ability of the individual tested. (See [11] for more details.).

The following two examples have a more mathematical character. Nevertheless, they do not lack their practical sense. Example 6 can be considered as the robust version of the stochastic approximation problem of finding the root of some regression function. Example 7 is a special case of 6 and deals with the problem of finding the γ-quantile having i.i.d. random variables.

Example 6. *Finding a root of the general regression quantile function*

Let observations at a point $x \in X \subseteq \mathbf{R}$ have a common distribution function G_x. Let the function $M(x) \equiv G_x^{-1}(\gamma) \equiv \inf\{t\,;\, G_x(t) \geqslant \gamma\}$ satisfy inequalities

$$M(x) < \alpha \quad \text{for} \quad x < x^*,$$
$$M(x) > \alpha \quad \text{for} \quad x > x^*$$

where x^* is a real number. The function M is called the regression quantile function. The problem is to find the point x^* provided the observations Z_x, distributed according to G_x, $x \in X$, are available.

We transform the variables Z_x to $Y_x \equiv I[Z_x \geqslant \alpha]$, where $I[A]$ denotes the characteristic function of the set A. Thus we have the 0–1 variables on various levels x, the probability of success being $F(x) = EI[Z_x \geqslant \alpha] = 1 - G_x(\alpha)$. For $x < x^*$ we clearly

have $F(x) < \gamma$, while for $x > x^*, F(x) \geqslant \gamma$ ($F(x) > \gamma$ under some mild assumptions). The problem of finding the solution of the equation $M(x) = \alpha$ having the observations Z_x is thus transformed to finding the solution of the equation $F(x) = \gamma$ using 0–1 observations Y_x.

Example 7. *Estimate of γ-quantile based on a random sample*

Let H denote a distribution function for which there exists exactly one γ^* such that $H(\gamma^*) = \gamma$. We desire to estimate recursively the value γ^* getting sequentially the variables X_i (one per time unit) distributed according to the function H.

If X is an arbitrary random variable distributed according to H then $Z_x \stackrel{\text{def}}{=} m(x) + X$ is distributed according to $G_x \stackrel{\text{def}}{=} H(y - m(x))$. Any increasing function can be taken in the place of m. Being at time i we put $Z_{x_i} = X_i + m(x_i)$, $i = 1, 2 \ldots$.

If H is strictly monotone then $G_x^{-1}(\gamma) = m(x) + \gamma^*$ and hence $G_x^{-1}(\gamma) < 0$ for $x < m^{-1}(-\gamma^*)$ and $G_x^{-1}(\gamma) > 0$ for $x > m^{-1}(-\gamma^*)$. Taking into account that m is a known function ($m(x) = x$ is usually considered) we transform the problem of finding the γ-quantile of H to the problem of finding a root of the regression quantile function. Using the transformation from Example 6 this problem can be further transformed to the problem of finding the solution $F(x) = \gamma$ having 0–1 observations $Y_{x_i} = I[Z_{x_i} \geqslant 0]$.

Although the examples introduced above have the common mathematical background, the algorithms that are going to be suggested are more suitable in some practical cases and less in other ones. In the sequel we will try to sort the algorithms together according to their mathematical character, giving some notes on their usefulness in specific applications.

General assumptions

The common feature of the methods introduced in this paper consists in defining the level (or levels) on which we perform our experiment at the next time unit. For all further developed methods and theorems of convergence we will suppose that our observations at time n depend on the past only through the levels on which they are performed. We also assume that the observations Y_x on the levels x are distributed according to the 0–1 law with a parameter $1 - F(x)$, where F is a real function having values between 0 and 1 ($0 \leqslant F(x) \leqslant 1$).

The majority of the methods cited below preserve their convergence properties also under more general assumptions concerning the distribution of Y_x and the behavior

of F. We do not mention these assumptions here, pointing out that the general results are formulated in the literature cited at the relevant places.

Up and down methods

The so called UP AND DOWN methods are characterized by a discrete set of levels at which the experiments are performed. We shall suppose that these levels form a set

$$L \equiv \{a + ld;\ l \text{ integer}\} \tag{1}$$

where a, d are given numbers. (Although in practice the set of levels can be of another type it can be transformed to (1) using a one to one transformation.)

Having an observation Y_{x_n} at time n we define the level of experiment at time $n+1$ as

$$\begin{aligned} x_{n+1} &= x_n + d \quad \text{if} \quad Y_{x_n} = 0, \\ x_{n+1} &= x_n - d \quad \text{if} \quad Y_{x_n} = 1. \end{aligned} \tag{2}$$

This property characterizes the UP AND DOWN methods.

One of the first papers concerning the recursive estimate of the *LD-50* variable was that of Dixon and Mood. They combine the UP AND DOWN principle with the maximum likelihood method. Their result consists in the assumption of normality of F. Since we introduce only non-parametric methods here we refer to [6] for the full algorithm and its numerical properties. A non-parametric UP AND DOWN method is, for example, that of Derman. Here we propose his algorithm (see [5]) in a little bit more general way.

Derman's approach

Assumptions P. Let there exist $x^* \in \mathbf{R}$ and $\gamma \in \left\langle \frac{1}{2}, 1 \right)$ such that

$$\begin{aligned} F(x) < \gamma \quad &\text{pro} \quad x < x^*, \\ F(x) > \gamma \quad &\text{pro} \quad x > x^*. \end{aligned}$$

Let $\liminf_{x \to \pm\infty} |F(x) - \gamma| > 0$.

Algorithm 1. Perform an observation at time n on the level x_n, where

$$
\begin{aligned}
&x_1 \in L \quad \text{arbitrary},\\
&x_{n+1} = x_n - d \quad \text{with probability} \quad \frac{1}{2\gamma} \quad \text{if} \quad Y_{x_n} = 1,\\
&x_{n+1} = x_n + d \quad \text{with probability} \quad 1 - \frac{1}{2\gamma} \quad \text{if} \quad Y_{x_n} = 1,\\
&x_{n+1} = x_n + d \quad \text{with probability} \quad 1 \quad \text{if} \quad Y_{x_n} = 0.
\end{aligned}
$$

(For $\gamma = \frac{1}{2}$ this algorithm coincides with (2).)

We define an estimate of x^* at time n as θ_n which is the element of

$$\arg\max_{i \in L} \sum_{k=1}^{\infty} I\{x_k = i\}. \tag{3}$$

We put θ_n to be the arithmetic mean of all the elements of (3) if this set has more than one element.

Assertion 1. Let the assumption P hold true. If $x^* \in L$, the relation

$$\theta_n \in \langle x^* - d, x^* + d\rangle \tag{4}$$

holds for all $n \geqslant N_0$, where N_0 is an a.s. finite random variable. If $x^* \notin L$ then $\theta_N \in \langle [x^*], [x^* + d]\rangle$, where $[x] \equiv \max\{y \in L;\, y \leqslant x\}$.

The proof of the assertion is given in [5] under the assumption that F is a distribution function. In [3] it was proved that the assertion remains valid also for $F(x) \underset{x \to \pm\infty}{\longrightarrow} \gamma$ provided this convergence is sufficiently slow.

The Derman method can be applied also in the case of $\gamma \in (0, \frac{1}{2})$. However, in this case the sequence $\{x_n\}_{n=1}^{\infty}$ is defined in the following manner:

$$
\begin{aligned}
&x_{n+1} = x_n + d \quad \text{with probability} \quad \frac{1}{2(1-\gamma)} \quad \text{if} \quad Y_{x_n} = 0,\\
&x_{n+1} = x_n - d \quad \text{with probability} \quad 1 - \frac{1}{2(1-\gamma)} \quad \text{if} \quad Y_{x_n} = 0,\\
&x_{n+1} = x_n - d \quad \text{with probability} \quad 1 \quad \text{if} \quad Y_{x_n} = 1.
\end{aligned}
$$

We should notice that the disadvantage of this method consists in the fact that $\limsup_{n \to \infty} |x_n - x^*| = \infty$. It means that the levels at which we perform observations can be arbitrarily far from the true value x^*. This fact should be taken into account, especially when the price of the experiment increases with the distance of the level from the value x^*.

Methods of the Robbins-Monro type

In this section we suppose that there exist $x^* \in \mathbf{R}$, $\gamma \in \mathbf{R}$ such that

$$\inf_{\varepsilon<|x-x^*|<1/\varepsilon}(F(x)-\gamma)(x-x^*) > 0,$$

for all $\varepsilon > 0$.

The Robbins-Monro method (RM) seems to be the method most frequently used for finding the root of the regression function. RM is the a recursive scheme, where the level of our experiment at time n is defined as the estimate of x^* at this time. In the simplest form the sequence of levels given by RM can be expressed as

$$x_{n+1} = x_n - a_n(Y_{x_n} - \gamma). \tag{5}$$

If $\sum a_n = \infty$ and $\sum a_n^2 < \infty$, then $x_n \to x^*$ almost surely (see [2] for proof).

An interesting insight into this method was given by Robbins, Lai [14]. They pointed out that the level x_{n+1} can be obtained as the root of the linear regression based on the observations $(x_i, Y_{x_i})_{i=1}^n$. Now we go briefly through their idea.

Suppose a theoretical model

$$Y_{x_i} = \beta x_i + \alpha + \varepsilon_i, \tag{L}$$

where ε_i are independent variables with $N(0,\sigma^2)$ distribution. If β is supposed to be known then we get the maximum likelihood estimate of the root of the equation $\beta x + \alpha = 0$ in the form

$$x_{n+1} = \bar{x}_n - \beta^{-1}\bar{y}_n, \tag{6}$$

where $\bar{x}_n = \frac{1}{n}\sum x_i$, $\bar{y}_n = \frac{1}{n}\sum(Y_{x_i} - \gamma)$. Setting $a_n = \frac{1}{\beta n}$, we get the identity between the sequences defined by the relations (5) and (6). In the paper of Robbins and Lai the parameter β (which is usually unknown in practice) is estimated using the maximum likelihood method again. Thus the recursive scheme suggested in their paper is of the form

$$x_{n+1} = x_n - \frac{Y_{x_n} - \gamma}{nb_n}, \tag{7}$$

where $b_n = b \vee (\tilde{\beta}_n \wedge B)$ for $n \geqslant n_0, b_n \in \langle b, B\rangle$ arbitrary for $n < n_0$,

$$0 < b < \beta < B, \quad n_0 = \inf\Big\{k;\ \sum_{i=1}^{k}(x_i - \bar{x}_k)^2 > 0\Big\} \quad \text{and} \quad \tilde{\beta}_n = \frac{\sum_{i=1}^{n} Y_{x_i}(x_i - \bar{x}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}.$$

Assertion 2. Let $F'(x^*) > 0$ and let F be continuously differentiable in a neighborhood of x^*. Then for the sequence $\{x_n\}_{n=1}^{\infty}$ defined in (7) the convergencies

$$\begin{aligned} x_n &\to x^* \quad \text{a.s.} \\ \sqrt{n}(x_n - x^*) &\xrightarrow{D} N\left(0, \gamma(1-\gamma)/\left(F'(x^*)\right)^2\right) \\ \tilde{\beta}_n &\to F'(x^*) \quad \text{a.s.} \end{aligned}$$

hold true.

See [15] for proof.

Using the results of Sacks [17] the asymptotic normality can be shown also for the sequences x_n defined in (5) with $a_n = \frac{a}{n}$, if $a > \frac{1}{2F'(x^*)}$. In this case the asymptotic variance is equal to $\frac{a^2\gamma(1-\gamma)}{2aF'(x^*)-1}$. This variance is minimized for $a = \left(F'(x^*)\right)^{-1}$. Thus the value b_n from (7) can be interpreted as the estimate of this optimal choice which is usually unknown in practice. The values b, B can be considered as an apriori knowledge concerning the parameter β.

Another approach to estimating the unknown value $\left(F'(x^*)\right)^{-1}$ is based on the differences of the function F. This method (in literature usually called adaptive) can be used also in the case when the apriori knowledge about b and B is not correct. From the results dealing with adaptive algorithms, that of Fabian [9] is well known:

$$x_{n+1} = x_n - \frac{d_n(Y_n - \gamma)}{n}, \tag{8}$$

where $Y_n = (Y_{x_n+c_n} + Y_{x_n-c_n})/2$, $d_n = \left((C_1 \log(n+1))^{-1} \vee A_n^{-1}\right) \wedge C_2 n^{\alpha}$,

$$A_n = (n-1)^{-1} \sum_{j=1}^{n-1} \frac{Y_{x_j+c_j} - Y_{x_j-c_j}}{2c_j},$$

$$0 < C_1 < C_2, \quad 0 < \alpha < \tfrac{1}{2}, \quad c_n = cn^{-\gamma}, \quad \gamma \in (\tfrac{1}{4}, \tfrac{1}{2}).$$

In practical realization of (8) we perform two observations at time n on the levels $x_n + c_n$, $x_n - c_n$.

Assertion 3. Let F have the second bounded derivative in a neighborhood of x^*, let $F'(x^*) > 0$. Then we get

$$\begin{aligned} x_n &\to x^* \quad \text{a.s.} \\ \sqrt{n}(x_n - x^*) &\xrightarrow{D} N(0, \gamma(1-\gamma)/\left(F'(x^*)\right)^2). \end{aligned}$$

See [9] for proof.

The methods of the RM type belong to the most frequently used recursive statistical approaches. One of their disadvantages in the case of estimating a variable *LD-50* may consist in the following fact. The experimental levels at which we can perform our observations usually form a discrete set (e.g. of the type of L). This fact contradicts the demands of the RM type algorithms. The transformation of the RM algorithm to the case of a discrete set of levels of observations can be found in [8]. Another disadvantage of these algorithms can appear when we need to know in advance the levels at which we shall perform our experiments. This situation arises e.g. in a preparation of psychological tests (see Example 5 & [11]). Fixing the number of items in a test (denote this number by k which is usually greater then 25) we need to combine all possibilities of answers to prepare the test on a computer. In the case of RM type methods this represents 2^k possibilities. On the other hand, the UP AND DOWN methods need only $\frac{1}{2}k(k+1)$ possibilities.

NON-PARAMETRIC METHODS BASED ON THE PARAMETRIC APPROACH

The methods of the RM type can be viewed as parameter estimates in the model (L). However, the linear fit is not the best what we can do in the case of 0–1 variables and a regression function which is a distribution one. Therefore other parameterizations were suggested using more suitable models. The convergence results were obtained also in these cases. The general idea of these methods can be expressed as follows: The function F is supposed hypothetically to be of the form $F(x) = H(x|\theta)$. Using observations $(x_i, Y_{x_i})_{i=1}^n$ we evaluate the maximum likelihood estimate $\tilde{\theta}_n$ of the parameter θ. For the level of our experiment at time $n+1$ we take the solution of the equation $\tilde{F}_n(x) = \gamma$, where $\tilde{F}_n(\cdot) = H(\cdot|\tilde{\theta}_n)$. This level is denoted by x_{n+1} as in the previous algorithms.

In the case of RM type methods we have $H(x|\theta) = \alpha + \beta x$, $\theta = (\alpha, \beta)$.

Wu in [18] suggested for the *LD-50* problem the following logistic parameterization of the function F : $H(x|\theta) = \left(1 + e^{-\lambda(x-\alpha)}\right)^{-1}$, $\lambda > 0$, $\theta = (\alpha, \lambda)$. In this case the maximum likelihood function is of the form

$$L(\lambda, \alpha | x_i, Y_{x_i}, i = 1, \ldots, n) = \prod_{i=1}^{n} H(x_i|\lambda, \alpha)^{Y_{x_i}} \left(1 - H(x_i|\lambda, \alpha)\right)^{(1-Y_{x_i})},$$

hence the likelihood equations are

$$\text{(9)} \qquad \begin{aligned} \sum_{i=1}^{n} H(x_i|\lambda, \alpha) &= \sum_{i=1}^{n} Y_{x_i}, \\ \sum_{i=1}^{n} x_i H(x_i|\lambda, \alpha) &= \sum_{i=1}^{n} x_i Y_{x_i}. \end{aligned}$$

Thus the estimate of the γ-quantile of F is defined as the value $\tilde{L}_\gamma = \tilde{\alpha} - \tilde{\lambda}^{-1} \log(1/\gamma - 1)$, where $\tilde{\alpha}$, $\tilde{\lambda}$ solve (9). This value defines also the level x_{n+1} on which we perform our experiment at time $n+1$. For practical purposes we should know under what condition there exists a solution of (9). Due to Silvapulle [19] there exists exactly one solution of (9) if

$$(x^+_{\min}, x^+_{\max}) \cap (x^-_{\min}, x^-_{\max}) \neq \emptyset$$

or

$$x^+_{\min} < x^-_{\min} = x^-_{\max} < x^+_{\max}$$

or

$$x^-_{\min} < x^+_{\min} = x^+_{\max} < x^-_{\max},$$

where

$$x^+_{\max} = \max\{x_i, Y_{x_i} = 1\}, \quad x^+_{\min} = \min\{x_i\,;\, Y_{x_i} = 1\},$$
$$x^-_{\max} = \max\{x_i\,;\, Y_{x_i} = 0\}, \quad x^-_{\min} = \min\{x_i\,;\, Y_{x_i} = 0\}.$$

If this condition is fulfilled for some n_0 then it is fulfilled also for all $n \geqslant n_0$. For $n < n_0$ we should use another method, e.g. of the RM type.

In [18] the convergence of this method is proved for λ given. For λ obtained as the solution of (9) the asymptotical efficiency of the method is proved provided the a.s. convergence holds true. In his paper Wu performs also some simulation studies according to which the previous algorithm gives better results then the RM method for small and medium n. See [18] for some numerical illustrations that exhibit better behavior of this method as compared with RM algorithm for small and medium sample sizes.

Methods based on a more general parametric model

Mukerjee in [12] suggested the approximation of the unknown distribution function (in his paper the regression function, generally) by the so called isotonic regression. The set of levels at which we perform our observations is the same as in the case of the UP AND DOWN methods, i.e. the experimental levels are taken from the set L defined in (1).

Let us introduce the following notation for the rest of our paper:
X_n ... levels, on which we performed our experiments up to time n,
$n(x)$ ... number of observations on the level x made up to time n,
$y_i(x)$ ... the value of the i-th observation on the level x $(i = 1, \ldots, n(x))$.

Isotonic regression based on the observations $(x_i, Y_{x_i})_{i=1}^n$ is defined as

$$\arg\min_{t \in RF} \sum_{x \in X_n} \sum_{i=1}^{n(x)} \big(y_i(x) - t(x)\big)^2, \tag{10}$$

where $RF = \{t : X_n \to R;\ t(x) \leqslant t(y),\ \text{if } x < y\}$. The recursive scheme based on the isotonic regression is described by the following algorithm:

Algorithm 2.

1) Choose $a < b \in \mathbb{R}$ arbitrarily and set $X_0 := L \cap \langle a, b\rangle$, $n := 0$. At every point of X_0 perform at least one observation.

2) Fit the values obtained up to time n by isotonic regression, i.e. find the element of (10) and denote it by t_n^*. (Notice that t_n^* is unique). Define

$$t_{\min}^{n,\gamma} \equiv \max\big\{\min X_n - d, \max\{x \in X_n\,;\, t_n^* < \gamma\}\big\},$$
$$t_{\max}^{n,\gamma} \equiv \min\big\{\max X_n + d, \min\{x \in X_n\,;\, t_n^* > \gamma\}\big\},$$

where we set $\max \emptyset = -\infty$, $\min \emptyset = \infty$.

Choose $\theta_n \in (t_{\min}^{n,\gamma}, t_{\max}^{n,\gamma})$ arbitrarily and perform experiments on the levels $[\theta_n]$, $[\theta_n + d]$, where $[y] = \max\{x \in L;\, x \leqslant y\}$. Set $n := n + 1$ and repeat step 2).

The general convergence theorem from the paper of Mukerjee can be rewritten for our purposes in the following way.

Assertion 4. Let there exist $x_m < x_M \in L$ such that

$$F(x) < \gamma \quad \text{for} \quad x \leqslant x_m,$$
$$F(x) > \gamma \quad \text{for} \quad x \geqslant x_M.$$

Then for θ_n from Algorithm 2 the relation

$$P[\theta_n \notin \langle x_m, x_M\rangle \text{ infinitely often}] = 0$$

is valid.

The proof is given in [12].

For completeness we should mention the algorithm for solving (10) which is, however, given e.g. in [1]. Here we would suggest other algorithms for finding the γ-quantile of the distribution function F whose character is similar to Mukerjee's algorithm but the problem (10) is simplified.

Instead of looking for the isotonic regression and for the interval $(t^{n,\gamma}_{\min}, t^{n,\gamma}_{\max})$ we estimate the γ-quantile directly as any element of

$$\arg\min_{\theta} \sum_{x \leqslant \theta} \sum_{i=1}^{n(x)} (1-\gamma) y_i(x) + \sum_{x \geqslant \theta} \sum_{i=1}^{n(x)} \gamma\big(1 - y_i(x)\big). \tag{11}$$

Recalling that $y_i(x)$ is equal to 0 or 1 we get that the set (11) is equal to

$$\arg\min_{\theta} \sum_{x \leqslant \theta} \sum_{i=1}^{n(x)} \big(y_i(x) - \gamma\big)^+ + \sum_{x \geqslant \theta} \sum_{i=1}^{n(x)} \big(y_i(x) - \gamma\big)^-, \tag{12}$$

where $z^+ = \max(0, z)$, $z^- = \max(0, -z)$. The following algorithm for finding any element of (11), or (12) is analogous to that one of Dupač [7] for finding the root of the quasiisotonic regression.

Algorithm 3. Set $x_1 := \min X_n$. Let $r_1 \in L$ be the smallest number such that

$$\sum_{\substack{x \in \langle x_1, r_1 \rangle \\ x \in X_n}} \sum_{i=1}^{n(x)} \big(y_i(x) - \gamma\big) \leqslant 0,$$

$r_2 > r_1$ the smallest number such that

$$\sum_{\substack{x \in \langle r_1 + d, r_2 \rangle \\ x \in X_n}} \sum_{i=1}^{n(x)} \big(y_i(x) - \gamma\big) \leqslant 0, \quad \text{etc.}$$

We take $\tilde{\theta} \in \langle -\infty, x_1)$ arbitrarily if r_1 does not exist. Otherwise we denote by r^* the last one from the r_k's defined and take $\tilde{\theta} \in (r^*, r^* + d)$ if $r^* < \max X_n$. If $r^* = \max X_n$, then select $\tilde{\theta} \in (r^*, \infty\rangle$. $\tilde{\theta}$ is an element of (11) or (12).

Let any $\tilde{\theta} \in (l_1, l_2)$ be an element of (11), where $l_1, l_2 \in L \cup \{\infty\}$, and let $\sum_{\substack{x \in (z, l_1) \\ x \in X_n}} \sum_{i=1}^{n(x)} \big(y_i(x) - \gamma\big) = 0$. Then any $\tilde{\theta} \in (z, z + d)$ is also an element of (11).

Denote

$$\theta^{n,\gamma}_{\min} = \max\big\{\min X_n - d;\ \inf\{\theta;\ \theta \text{ is an element of (11)}\}\big\}$$

and

$$\theta^{n,\gamma}_{\max} = \min\big\{\max X_n + d;\ \sup\{\theta;\ \theta \text{ is an element of (11)}\}\big\}.$$

The relation between $\theta^{n,\gamma}_{\max(\min)}$ and $t^{n,\gamma}_{\max(\min)}$ is expressed in the following assertion.

Assertion 5. $\theta_{\min}^{n,\gamma} \geqslant t_{\min}^{n,\gamma}$, $\theta_{\max}^{n,\gamma} \leqslant t_{\max}^{n,\gamma}$.

Proof. We denote $\theta_{\min} = \theta_{\min}^{n,\gamma}$ and $t_{\min} = t_{\min}^{n,\gamma}$ for simplicity. We prove the first inequality, the second could be proved similarly. We suppose that $t_{\min} > \min X_n - d$, otherwise the assertion is clear. Let $\theta_{\min} < t_{\min}$.

Denote by t^* the solution of (10). There must exist a natural number c such that

$$t^*(\theta_{\min} + d) = t^*(\theta_{\min} + 2d) = \ldots = t^*(\theta_{\min} + cd).$$

Define

$$M = \{\theta_{\min} + d, \theta_{\min} + 2d, \ldots, \theta_{\min} + cd\}$$

and

$$\tilde{t}(x) = \min\{\gamma, t^*(\theta_{\min} + (c+1)d)\} \quad \text{for} \quad x \in M,$$
$$\tilde{t}(x) = t^*(x) \quad \text{for} \quad x \in X_n \backslash M.$$

It is obvious that $\sum_{x \in M} \sum_{i=1}^{n(x)} (y_i(x) - k)^2$ is a decreasing function in k for $k \leqslant \frac{1}{n(M)} \sum_{x \in M} \sum_{i=1}^{n(x)} y_i(x)$, where $n(M) = \sum_{x \in M} \sum_{i=1}^{n(x)} 1$ is the number of observations performed on the levels from M. Further, we get $\frac{1}{n(M)} \sum_{x \in M} \sum_{i=1}^{n(x)} y_i(x) \geqslant \gamma$ using the inequality $\sum_{x \in M} \sum_{i=1}^{n(x)} y_i(x) - \gamma \geqslant 0$ which follows from the definition of $\theta_{\min}$ by virtue of (12). This fact implies that the function $\sum_{x \in M} \sum_{i=1}^{n(x)} (y_i(x) - k)^2$ is decreasing for $k \leqslant \gamma$.

On the set M the values $\tilde{t}$, t^* are constant, less than γ and $\tilde{t} > t^*$. This contradicts the fact that t^* is an element of (10). □

The next result is an immediate corollary of the previous assertion and Assertion 4.

Assertion 6. Let the assumptions of Assertion 4 be valid. Replacing $t_{\min}^{n,\gamma}$ by $\theta_{\min}^{n,\gamma}$ and $t_{\max}^{n,\gamma}$ by $\theta_{\max}^{n,\gamma}$ in Algorithm 2 we get for θ_n from this algorithm that

$$P[\theta_n \notin \langle x_m, x_M \rangle \text{ for infinite number of } n] = 0.$$

Algorithm 2 thus can be used taking the estimate of the γ-quantile based on solving (11) or (12). These problems are easier to solve than the isotonic regression problem.

The algorithms studied in the last two sections are particularly useful in such situations when some observations are given in advance on different levels without any order.

At the end we give several numerical results that demonstrate the behavior of the algorithms introduced. Table 1 shows the results when the median of the $N(0,1)$ distribution was to be found. In Table 2 we consider the problem of finding 70 % quantile of χ^2 (8) distribution which is equal to 9.524. The methods used were the Derman method using Algorithm 1, Robbins-Monro method (according to (5) with $a_n = \frac{a}{n}$), adaptive Robbins-Monro method according to (8). In this method the constants d_n were replaced by $d'_n = (A_n^{-1} \vee r_1) \wedge r_2$, where r_1, r_2 are chosen constants and A_n has the same meaning as in the definition of d_n. Further, the isotonic regression approach using algorithm 2 and Robbins-Lai approach using formula (7) were employed. Each method was applied from time 0 to **Stoptime** giving one sample trajectory of the algorithm. In order to obtain more knowledge about the asymptotic properties of our procedures we took **T** such sample paths for each of the procedures. The value of **T** is given in the first column of the numerical tables. We do not introduce the whole history of the processes but only their values at the times that are given in the second column. In the third column the values of specified procedures are computed at the time moments considered. These values are taken from the last path of our **T** samples. In the fourth column the averages through the **T** samples are given. Finally, the sample variances of our procedures at specified time moments are given in the last column. We do not comment the results because they depend very heavily on the parameters of each of our procedures. However, the results probably give some ideas concerning the procedures behavior. For more information the time consumption per one path simulation is given (results were computed on IBM XT compatible without coprocessor). This information can be misleading in the case of the Derman procedure because type of random number generator other than for the other four methods had to be used. We use the Stochastic Approximation program system for our computations (see Charamza [4]). Using this system a lot of simulations under different initial conditions and different parameter values can be easily obtained.

TABLE 1

Derman procedure.

$d = 0.1000$, StopTime = 100, Time consumption = 13.509000 s

T	time	estimate	average	variance
20	0	2.000000	2.0000	0.0000
20	10	1.500000	1.5225	0.0322
20	20	0.850000	0.9148	0.1191
20	30	0.600000	0.5946	0.0985
20	40	0.600000	0.3423	0.0955
20	50	−0.050000	0.2375	0.1163
20	60	−0.050000	0.1375	0.0665
20	70	0.000000	0.1200	0.0430
20	80	−0.050000	0.0525	0.0583
20	90	−0.100000	0.0175	0.0419
20	100	0.000000	0.0275	0.0385

Procedure Robbins-Monroe.

$a = -3.0000$, StopTime = 100, Time consumption = 6.838000 s

T	time	estimate	average	variance
20	0	2.000000	2.0000	0.0000
20	10	0.314881	0.1136	0.0986
20	20	0.487888	0.1045	0.0663
20	30	0.465465	0.0536	0.0514
20	40	0.217214	0.0485	0.0405
20	50	0.158326	0.0470	0.0335
20	60	0.104783	0.0518	0.0261
20	70	0.194576	0.0623	0.0272
20	80	0.153601	0.0527	0.0197
20	90	0.049032	0.0497	0.0156
20	100	0.018173	0.0355	0.0126

Adaptive Robbins-Monro procedure.

$c = 1.0000$, gamma = 0.3000, $r_1 = 0.1000$, $r_2 = 5.0000$, StopTime = 100, Time consumption = 13.534000 s

T	time	estimate	average	variance
20	0	2.000000	2.0000	0.0000
20	10	−0.197529	0.0376	0.1476

20	20	−0.220077	0.0828	0.0666
20	30	−0.137035	0.0536	0.0410
20	40	−0.080158	0.0433	0.0230
20	50	−0.005332	0.0519	0.0186
20	60	0.070421	0.0569	0.0165
20	70	0.013817	0.0379	0.0118
20	80	−0.039966	0.0253	0.0107
20	90	−0.103091	0.0239	0.0080
20	100	−0.103091	0.0202	0.0078

Isotonic regression (see Algorithm 2).

$d = 0.1000$, StopTime = 100, Time consumption = 12.883000 s

T	time	estimate	average	variance
20	0	2.000000	2.0000	0.0000
20	10	1.250000	1.2050	0.0205
20	20	0.450000	0.6000	0.0563
20	30	0.150000	0.3550	0.0752
20	40	0.150000	0.2000	0.0647
20	50	0.150000	0.1250	0.0504
20	60	0.150000	0.1250	0.0251
20	70	0.150000	0.0500	0.0168
20	80	0.150000	0.0350	0.0192
20	90	0.150000	0.0450	0.0173
20	100	0.050000	0.0150	0.0150

Procedure of Robbins and Lai.

$b = 0.1$, $B = 5.0000$, StopTime = 100, Time consumption = 8.271500 s

T	time	estimate	average	variance
20	0	2.000000	2.0000	0.0000
20	10	−0.572960	−0.0774	0.1535
20	20	−0.393169	−0.0151	0.0801
20	30	−0.295483	−0.0614	0.0416
20	40	−0.199799	−0.0190	0.0388
20	50	−0.193787	−0.0140	0.0415
20	60	−0.246730	−0.0036	0.0308
20	70	−0.102382	0.0165	0.0238
20	80	−0.019223	0.0230	0.0185
20	90	−0.094940	0.0270	0.0181
20	100	−0.125221	0.0124	0.0161

TABLE 2

Derman procedure.

$d = 0.4000$, StopTime = 200, Time consumption = 161.932000 s

T	time	estimate	average	variance
10	0	2.000000	2.0000	0.0000
10	20	6.000000	5.7467	1.8744
10	40	6.200000	7.7200	1.5307
10	60	9.200000	9.0800	1.9307
10	80	9.200000	8.9400	1.4316
10	100	8.200000	8.8600	1.3249
10	120	8.200000	8.8000	1.3244
10	140	9.200000	9.3000	0.3667
10	160	9.200000	9.2800	0.3129
10	180	9.200000	9.3000	0.2600
10	200	9.200000	9.4600	0.4804

Procedure Robbins-Monroe.

$a = -5.0000$, StopTime = 200, Time consumption = 41.689000 s

T	time	estimate	average	variance
10	0	2.000000	2.0000	0.0000
10	20	9.271491	8.5825	0.8635
10	40	9.917205	8.9442	0.4873
10	60	10.027956	9.1096	0.3826
10	80	9.808527	9.1442	0.3134
10	100	9.748349	9.2024	0.2568
10	120	9.787566	9.2491	0.2426
10	140	9.829480	9.2465	0.2445
10	160	9.799907	9.2425	0.2217
10	180	9.831373	9.2929	0.1949
10	200	9.750808	9.2982	0.1624

Adaptive Robbins Monro.

$c = 1.0000$, gamma = 0.3000, $r_1 = 0.1000$, $r_2 = 5.0000$, StopTime = 200, Time consumption = 84.525000 s

T	time	estimate	average	variance
10	0	2.000000	2.0000	0.0000
10	20	8.408059	9.6082	1.3243

10	40	8.380440	9.5545	0.8003
10	60	8.353868	9.5636	0.5391
10	80	8.500834	9.4958	0.3527
10	100	8.521206	9.4720	0.2723
10	120	8.553610	9.4962	0.2355
10	140	8.641721	9.4900	0.1562
10	160	8.722310	9.5018	0.1245
10	180	8.783905	9.4855	0.0973
10	200	8.754266	9.4878	0.1151

Isotonic regression.

$d = 0.4000$, StopTime = 200, Time consumption = 83.092000 s

T	time	estimate	average	variance
10	0	2.000000	2.0000	0.0000
10	20	7.800000	7.6000	0.6133
10	40	9.400000	9.1600	0.4693
10	60	10.600000	9.5600	0.7538
10	80	9.800000	9.5200	0.2862
10	100	9.800000	9.3600	0.3716
10	120	9.000000	9.3200	0.2773
10	140	9.400000	9.4000	0.2844
10	160	9.400000	9.4000	0.1778
10	180	9.400000	9.4800	0.1351
10	200	9.400000	9.5600	0.2560

Robbins and Lai procedure.

$b = 0.1$, $B = 5.0000$, StopTime = 200, Time consumption = 43.693000 s

T	time	estimate	average	variance
10	0	2.000000	2.0000	0.0000
10	20	9.349952	9.6220	1.3684
10	40	10.330679	9.7979	0.4026
10	60	10.352507	9.8537	0.3575
10	80	9.936621	9.7053	0.2875
10	100	9.830108	9.7168	0.2212
10	120	9.900507	9.7468	0.2218
10	140	9.977772	9.6548	0.2195
10	160	9.865312	9.5860	0.2169
10	180	9.921364	9.6457	0.1781
10	200	9.781967	9.6193	0.1593

Final remarks

The algorithms introduced here do not cover all possibilities for solving the problems of *LD-50* type. Other algorithms of the RM type were introduced (see [13]) even for the case of dependent observations (see [10]). By their modifications we could gain many other schemes. We should mention at least one reference concerning the non-recursive methods for solving the problem studied above. Several methods of this type were studied intensively in [16].

References

[1] *Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D.*: Statistical Inference Under Order Restrictions, Wiley, London, 1972.

[2] *Blum, J.R.*: Multidimensional stochastic approximation procedures, AMS *25* (1954), 737–744.

[3] *Charamza, P.*: Stochastic approximation on the lattice, Diploma work, Faculty of mathematics and physics, Charles University Prague, 1984. (In Czech.)

[4] *Charamza, P.*: Software for stochastic approximation, Proceedings of the conference ROBUST 1990.

[5] *Derman, C.*: Non-parametric up-and-down experimentation, AMS *28* (1957), 795–797.

[6] *Dixon, Mood*: A method for obtaining and analyzing sensitivity data, JASA *43* (1948), 109–126.

[7] *Dupač, V.*: Quasiisotonic regression and stochastic approximation, Metrika *34* (1987), 117–123.

[8] *Dupač, V., Henkenrath, U.*: On integer stochastic approximation, Aplikace matematiky *29* (1984), 372–383.

[9] *Fabián, V.*: On asymptotic normality in stochastic approximation, AMS *39* (1968), 1327–1332.

[10] *Holst, U.*: Recursive estimation of quantiles, Contributions to probability and statistics in honor of Gunnar Blom, Univ. Lund, 1985, pp. 179–188.

[11] *Lord, F.M.*: Tailored testing. An application of stochastic approximation, Journal of the American Statistical Association *66* (1971), 707–711.

[12] *Mukerjee, H.G.*: A stochastic approximation by observation on a discrete lattice using isotonic regression, AMS *9* (1981), 1020–1025.

[13] *Nevelson, M.B.*: On the properties of the recursive estimates for a functional of an unknown distribution function, Limit Theorems of Probability Theory (P. Revezs, eds.), Amsterodam, 1975, pp. 227–252.

[14] *Robbins H., Lai T.L.*: Adaptive design and stochastic approximation, AS *7* (1979), 1196–1221.

[15] *Robbins H., Lai T.L.*: Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes., Z. Wahrsch. Verw. Gebiete *56* (1981), 329–360.

[16] *Roth, Josífko, Malý, Trčka*: Statistical methods in experimental medicine, Státní zdravotnické nakladatelství, Praha, 1962, pp. 592. (In Czech.)

[17] *Sacks, J.*: Asymptotic distribution of stochastic approximation procedures, AMS *29* (1958), 373–405.

[18] *Wu, C.F.J.*: Efficient sequential designs with binary data, Journal of the American Statistical Association *80* (1985), 974–984.

[19] *Silvapulle, M.J.*: On the existence of maximum likelihood estimators for the binomial response problem, Journal of the Royal Statistical Society, Ser. B *43* (1981), 310–313.

REKURZIVNÍ ODHADY KVANTILŮ ZALOŽENÉ NA 0–1 POZOROVÁNÍCH

PAVEL CHARAMZA

V práci jsou uvedeny některé rekurzivní procedury pro nalezení veličiny *LD-50* máme-li k dispozici nula-jedničková pozorování. Tyto metody mohou být obecněji použity pro odhad γ-kvantilu neznámé distribuční funkce. Kromě standardních metod typu Robbins-Monro jsou uvedeny i přístupem odlišné metody. Jsou rovněž zmíněny praktické možnosti použití zvolených metod.

Author's address: RNDr. *Pavel Charamza*, CSc., KPMS MFF UK, Sokolovská 83, 186 00 Praha 8.