Journal of Statistics: Advances in Theory and Applications
Volume 2, Number 2, 2009, Pages 185-199

REDUCED GRADIENT METHOD FOR MINIMAX ESTIMATION OF A BOUNDED POISSON MEAN

A. EL MOUATASIM and A. AL-HOSSAIN

Department of Mathematics
Faculty of Sciences
Jazan University
B.P. 2097, Jazan
Saudi Arabia
e-mail: a.mouatasim@gmail.com
aalhossain@jazanu.edu.sa

Abstract

This paper is concerned with minimax estimation of a bounded mean Poisson given the prior information that the mean lies in a bounded domain, and using the information normalized squared error loss. It would only be necessary to find the correct corresponding Bayes estimator, and we present the algorithm of reduced gradient via stochastic perturbation for solving this statistical problem.

1. Introduction

Two important problems in statistical inferences are estimation and tests of hypotheses. One type of estimation, namely, point estimation, is to be the subject of this work.

Point estimation admits two problems: the first, to devise some means of obtaining a statistic to use as an estimator; the second, to select criteria

2009 Mathematics Subject Classification: 62F15, 62F10, 62C20, 90C26.

Keywords and phrases: Bayes risk, loss estimation, minimax risk, global optimization.

Received September 30, 2009

and techniques to define and find a "best" estimator among many possible estimators several method. There are several methods of obtaining point estimators of parameters, among of these are (i) the Bayes method, (ii) maximum likelihood method (see, for instance, [4]). A Bayes estimator is given as the mean of the posterior or from the decision theoretical viewpoint as an estimator having smallest average risk (see, for instance, [7], [10], and [11]).

Usually, for all these criteria, the estimators are characterized as the global minima of non-convex functionals and the usual difficulties of non-convex optimization have been observed, when numerical procedures are introduced. In this framework, recent works have shown that suitable stochastic perturbations of usual reduced gradient method lead to robust method of global optimization (see, for instant, [1]).

We shall be concerned here with the estimation of a Poisson mean. Since, the Poisson distribution provides a realistic model for many random phenomena, and certain random experiments involving counts of happening in time (or length, space, area, volume, etc.) can be realistically modelled by assuming a Poisson distribution. Such a count might be:

(a) the number of fatal traffic accidents per week in a given state,

(b) the number of radioactive particle emissions per unit of time,

(c) the number of telephone calls per hour coming into the switchboard of a large business,

(d) the number of meteorites that collide with a test satellite during a single orbit,

(e) the number of organisms per unit volume of same fluid,

(f) the number of defects per unit of same material,

(g) the number of flaws per unit length of same wire.

We are given a random variable **X** with Poisson distribution; that is, the probability that $\mathrm{X} = \mathbf{x}$, denoted by $P(\mathrm{X} = \mathbf{x}) = \mathrm{e}^{-\lambda} \frac{\lambda^{\mathbf{x}}}{\mathbf{x}!}$ for $\mathbf{x} = 0, 1, ..$ and $= 0,$ otherwise.

The mean of the distribution is λ. The parameter λ in the Poisson distribution is usually unknown. Using minimax estimation for a bounded mean, this statistical problem was first elaborated in a series of papers (see, for instances, [6] and [9]). This paper is concerned with minimax estimation of λ given the prior information that the Poisson mean λ lies in a bounded domain Λ in $\mathbb{R}_+ = [0, \infty)$, and using the information normalized squared error loss function

$$L(\delta, \lambda) = \frac{(\delta - \lambda)^2}{\lambda}. \tag{1}$$

This loss function is a commonly used compromise between mathematical tractability and statistical relative (see, for instance, [8]). However, the method would remain the same for estimation of different bounded parameters and different loss function. It would only be necessary to find the correct corresponding Bayes estimation by using those other densities and other loss functions.

The expected error or risk of an estimator δ, when λ is true, $R_\delta(\lambda) = E[L(\delta(\mathrm{X}), \lambda)]$. As a measure of the information in the experiment, we study the minimax risk

$$\rho(\Lambda) = \min_{\delta} \max_{\Lambda} R_\delta(\lambda). \tag{2}$$

Unfortunately, exact analytic description of the minimax rule and risk is generally intractable, since the minimax rule is a Bayes estimator for a prior π^*, the least favorable distribution, which is necessary concentrated on a meager set of complicated form. For simplest cases, in this paper, we interested with π^* consists of a small number of points.

2. Notation and Assumption

We assume throughout that T is relatively open in $\mathbb{R}^+$. Since, the continuity of risk functions ensures that $\rho(\Lambda) = \rho(\overline{\Lambda})$, we may and shall by convention choose Λ, so that, $\Lambda = \operatorname{int} \overline{\Lambda}$.

The problem of estimation, as it shall be considered herein, is loosely defined as follows: Assume that some characteristic of the elements in a population can be represented by a random variable X, whose density is $fx(., \lambda) = f(., \lambda)$, where the form of the density is assumed known except that, it contains an unknown parameter λ.

Further, assume that the value of x of a random sample X from $f(., \lambda)$ can be observed. On the basis of the observed sample values x, it is desired to estimate the value of the unknown parameter λ, of the unknown parameter. This estimation can be made in two ways. The first, called point estimation, is to let the value of same statistic, say $\delta(x)$, represent or estimate, the unknown λ; such a statistic $\delta(x)$ is called a point estimator. The second, called interval estimation, is to define two statistics, say $\delta_1(x)$ and $\delta_2(x)$, where $\delta_1(x) < \delta_2(x)$, so that $(\delta_1(x), \delta_2(x))$ constitutes an interval, for which the probability can be determined that it contains the unknown λ.

Consider estimating λ, denote δ an estimate of λ.

The normalized squared loss function, denoted by

$$L(\delta, \lambda) = \frac{(\delta - \lambda)^2}{\delta}.$$

The risk function, denoted by $R_\delta(\lambda)$, of an estimator $\Delta = \delta(X)$ is defined to be

$$R_\delta(\lambda) = E[L(\Delta, \lambda)],$$

the risk function is the average loss.

Remark 1. If $L(\delta, \lambda) = (\delta - \lambda)^2$, then $R_\delta(\lambda) = E[(\Delta - \lambda)^2]$ is the **mean square** value of the estimation error (see, for instance, [12]) and we have shown, it minimum if $\lambda = E[\Delta]$.

Definition 1. An estimator δ^* is defined to be a *minimax estimator*, if and only if

$$\sup_{\lambda} R_{\delta^*}(\lambda) \le \sup_{\lambda} R_{\delta}(\lambda),$$

for every estimator δ.

Throughout the paper, we restrict attention to the class of estimators

$$D = D(\Lambda) = \{\delta(x) : x = 0 \text{ implies } \delta(x) = 0\}, \tag{3}$$

since estimators not in D are easily seen to have infinite maximum risk.

Definition 2. For a (prior) probability distribution $\pi(d\lambda)$, define the *integrated risk* $r(\delta, \pi)$, and the *Bayes risk* $r(\pi)$ by

$$r(\delta, \pi) = \int R_{\delta}(\lambda)\pi(d\lambda), \quad r(\pi) = \inf_{\delta \in D} r(\delta, \pi). \tag{4}$$

Definition 3. Let $\pi^*(X)$ denote the collection of probability measures supported in X. According to the minimax theorem,

$$\rho(\Lambda) = \inf_{D} \sup_{\Lambda} R_{\delta}(\lambda) = \sup_{\pi^*(\Lambda)} \inf_{D} r(\delta, \pi) = \sup_{\pi^*(\Lambda)} r(\pi). \tag{5}$$

A prior distribution attaining the supremum is called *least favorable* for Λ.

Definition 4. The Bayes estimator of λ, denoted by $\Delta^* = \delta^*(X)$, is defined to be that estimator with smallest Bayes risk, or the Bayes estimator of λ is the estimator δ^* satisfying

$$r(\delta^*, \pi) \le r(\delta, \pi),$$

for every other estimator $\Delta = \delta(X)$ of λ.

3. Bayes Estimator

In this section, we will see how Bayes estimation can sometimes be used to find a minimax estimator. Our objective has been to minimize risk, but since risk depended on the parameter, we were unable to find one estimator that had smaller risk than all others for all parameter values. Minimax circumvented such difficulty by replacing the risk function by its maximum value, and then seeking that estimator, which minimized such maximum value. Another way of getting around the difficulty arising from attempting to uniformly minimize risk is to replace the risk function by the area under the risk function, and to seek that estimator, which has the least area under its risk function.

Consider an interval $\overline{\Lambda} = [m_1, m_2]$ with $0 \leq m_1 < m_2 < \infty$. Since, the least favorable distribution π_{m_1, m_2} is unique, and analyticity considerations imply that it is supported on a finite number of points in $[m_1, m_2]$. Denote by $\epsilon_{\{\alpha\}}$, the probability measure concentrated at $\lambda = \alpha$. Let $\mathcal{F}_k[m_1, m_2]$ be the class of distributions of the form

$$\pi = \pi(a, \alpha) = \sum_{i=1}^{k} a_i \epsilon_{\{\alpha_i\}}, \tag{6}$$

where $\{\alpha_i\} \subset [m_1, m_2]$ and $\{a_i\}$ are probability masses summing to 1. The least favorable distribution π_{m_1, m_2} belongs to $\mathcal{F}_k[m_1, m_2]$ for k large enough. The optimization problem is then to choose $a_1, \ldots, a_k, \alpha_1, \ldots, \alpha_k$, constrained as above, so as to maximize

$$r_k(a, \alpha) = r(\pi(a, \alpha)) = \sum_{i=1}^{k} a_i R_{\delta_\pi}(\alpha_i). \tag{7}$$

Lemma 1. *If* δ *is minimax, then* $\delta(0) = 0$.

Proof. Let $F(\lambda) = R_{\delta_\pi}(\lambda)$, the function F is continuous on 0, then

$$R_{\delta_\pi}(0) = \lim_{\epsilon \to 0} R_{\delta_\pi}(\epsilon), \tag{8}$$

by the Equation (13), we have

$$\lim_{\epsilon \to 0} \frac{\partial R_{\delta_{\pi^*}}}{\partial \delta_{\pi^*}}(\epsilon) \,|_{x=0} = \frac{\partial R_{\delta_{\pi^*}}}{\partial \delta_{\pi^*}}(m) \,|_{x=0},$$

then

$$\lim_{\epsilon \to 0} \frac{\epsilon - \delta(0)}{\epsilon} = \frac{m - \delta(0)}{m},$$

so $\delta(0) = 0$.

Theorem 1. *The Bayes rule* δ_π *associated to such a prior is given by*

$$\delta_\pi(x) = \frac{1}{E[\lambda^{-1}|X = x]} = \frac{\sum_{i=1}^{k} a_i \alpha_i^x e^{-\alpha_i}}{\sum_{i=1}^{k} a_i \alpha_i^{x-1} e^{-\alpha_i}}, \tag{9}$$

with the convention (3) *requiring that* $\delta_{\pi(0)} = 0$.

Proof. π^* is maximum if

$$\sum_{i=1}^{k} a_i \frac{\partial R_{\delta_{\pi^*}}(\alpha_i)}{\partial \delta_{\pi^*}} = 0, \qquad \text{for } x \neq 0,$$

where

$$R_{\delta_{\pi^*}}(\alpha_i) = \sum_{x=0}^{\infty} \frac{(\delta(x) - \alpha_i)^2}{\alpha_i} \frac{\alpha_i^x}{x!} \mathrm{e}^{-\alpha_i},$$

then

$$\sum_{i=1}^{k} a_i (\alpha_i - \delta(x)) \frac{\alpha_i^{x-1}}{x!} e^{-\alpha_i} = 0, \quad \text{for} \quad x \neq 0,$$

so we have

$$\delta_{\pi^*}(x) = \frac{\sum_{i=1}^{k} a_i \alpha_i^x e^{-\alpha_i}}{\sum_{i=1}^{k} a_i \alpha_i^{x-1} e^{-\alpha_i}}, \quad x \neq 0.$$

For identify the maximum, we recall an important and familiar criterion:

Lemma 2. *If the support of a prior* $\pi(d\lambda)$ *is contained in the set, at which* $R_{\delta_\pi}(\lambda)$ *achieves its maximum on* $[m_1, m_2]$, *then* π *is least favorable and* δ_π *is minimax.*

Analytical descriptions for $[0, m]$, m is small. When m is sufficient small, it is plausible that the least favorable distribution $\pi_m := \pi_{0,m}$, on $[0, m]$ would be given by a pint mass at m.

Theorem 1. *If* $\Delta^* = \delta^*(X)$ *is a Bayes estimator having constant risk, that is,* $R_{\delta^*}(\lambda) \equiv$ *constant, then* Δ^* *is a minimax estimator.*

Proof. Let $g^*(.)$ be the prior density corresponding to the Bayes estimator $\delta^*(.)$.

$$\begin{aligned}\sup_{\lambda\in\Lambda} R_{\delta^*}(\lambda) &= \text{constant}\\ &= R_{\delta^*}(\lambda)\\ &= \int_\Lambda R_{\delta^*}(\lambda) g^*(\lambda) d\lambda\\ &\le \int_\Lambda R_\delta(\lambda) g^*(\lambda) d\lambda\\ &\le \sup_{\lambda\in\Lambda} R_\delta(\lambda),\end{aligned}$$

for any estimator $\delta(.)$.

4. Global Optimization Problem

We shall be concerned here with the estimation of a bounded Poisson mean with respect to a normalized squared error loss. However, the method would remain the same for estimation of different bounded parameters and different loss functions. It would only be necessary to find the correct corresponding Bayes estimator by using those other densities and other loss functions (see, for instance, [5] and [6]). The problem can be stated more precisely as follows.

Let us consider the statistical problem of making an inference about the unknown parameter λ, given an observation of X.

A solution consists of an estimator or decision procedure δ, which is a measurable function from $N \to [0, m] = \Lambda$, where N denotes the space of nonnegative integers.

D denotes the space of decision procedures. A risk function $R_\delta(\lambda)$, characterizes the performance of a decision procedure δ for each λ value of the parameter λ. The risk function is usually defined in terms of an underlying loss function $L(\delta, \lambda)$.

A loss function maps $D \times \Lambda \to \mathbb{R}^+ \cup \{0\}$ defines the cost of estimating δ, when λ is the true value of the parameter. Henceforth, we will assume that

$$L(\delta, \lambda) = \frac{(\delta - \lambda)^2}{\lambda}.$$

The risk of an estimator δ, when λ is true, $R_\delta(\lambda)$ is then the average loss incurred from using δ, that is, $R_\delta(\lambda) = E[L(\delta(X), \lambda)]$ and since $E[L(\delta(X), \lambda)] = \sum_x L(\delta(X), \lambda)P(X = x)$ see [12], then

$$R_\delta(\lambda) = \sum_{x=0}^{\infty} \frac{(\delta(x) - \lambda)^2}{\lambda} \frac{\lambda^x}{x!} e^{-\lambda}. \tag{10}$$

By Lemma 1

$$R_\delta(\lambda) = \lambda e^{-\lambda} + \sum_{x=1}^{\infty} \frac{(\delta(x) - \lambda)^2}{\lambda} \frac{\lambda^{x-1}}{x!} e^{-\lambda}. \tag{11}$$

A minimax estimator for the above problem is δ^*, if

$$\sup_{0 \le \lambda \le m} R_{\delta^*}(\lambda) \le \sup_{0 \le \lambda \le m} R_\delta(\lambda), \text{ for all } \delta \in D. \tag{12}$$

Minimax problems are often solved by considering the corresponding Bayes problems. A distribution or prior probability measure π is specified on the parameter space Λ, and the relative performance of a procedure δ is specified by its Bayes risk.

It follows from Ghosh's results [3] that a least favorable prior $[0, m] = \Lambda$, will put mass on at most a finite number of points.

More precisely, as shown by Johnstone and MacGibbon [8], the "least favorable" prior $\pi(d\lambda)$ is of the form $\sum_{i=1}^{k} a_i \epsilon_{\{\alpha_i\}}$ $(k < \infty)$. In addition, the Bayes rule δ_{π^*} is an "equalizer" rule; that is,

$$R_{\delta_{\pi^*}}(\alpha_i) = R_{\delta_{\pi^*}}(\alpha_j)\ \forall i,\ j = 1, \ldots, k. \tag{13}$$

It is known that $\alpha_1 = 0$ and $\alpha_k = m$ (see [8]).

Since, the objective function (7) of optimization problem is nonconvex,

$$\text{Maximize}: \sum_{i=1}^{k} a_i R_{\delta_\pi}(\alpha_i),$$

$$\text{Subject to}: \sum_{i=1}^{k} a_i = 1,$$

$$0 \leq a_i, \quad i = 1, 2, \ldots, k,$$

$$0 \leq \alpha_i \leq m, \quad i = 1, \ldots, k, \tag{14}$$

we will restrict ourselves to the problem, when $k = 3$.

Thus, our problem is reduced to the following: for a suitable $m_2 > m_1 \simeq 1.27$, and for each $m \in (m_1, m_2]$, find the a_1, a_2, and λ^* satisfying $0 \leq a_1 \leq 1, 0 \leq a_2 \leq 1$, and $0 < \lambda^* < m$ such that $\pi^* = a_1 \epsilon_{\{0\}} + a_2 \epsilon_{\{\lambda^*\}} + (1 - a_1 - a_2)\epsilon_{\{m\}}$ is least favorable. In this case, the problem reduces to a global optimization problem with equality linear constraint.

$$\text{Maximize}: a_1 R_{\delta_\pi}(0) + a_2 R_{\delta_\pi}(\alpha_2) + a_3 R_{\delta_\pi}(m),$$

$$\text{Subject to}: a_1 + a_2 + a_3 = 1,$$

$$0 \leq a_i, \quad i = 1, 2, 3,$$

$$0 \leq \alpha_2 \leq m. \tag{15}$$

By Equation (10), we have

$$R_{\delta_{\pi^*}}(0) = (\delta(1))^2,$$

$$R_{\delta_{\pi^*}}(\alpha_2) = \alpha_2 \mathrm{e}^{-\alpha_2} + \sum_{x=1}^{\infty} (\delta(x) - \alpha_2)^2 \frac{\alpha_2^{x-1}}{x!} \mathrm{e}^{-\alpha_2},$$

$$R_{\delta_{\pi^*}}(m) = m\mathrm{e}^{-m} + \sum_{x=1}^{\infty} (\delta(x) - m)^2 \frac{m^{x-1}}{x!} \mathrm{e}^{-m},$$

then the problem (15) is nonconvex optimization problem with 5 unknown $(\theta, a_1, a_2, a_3, \alpha_2)$ and 3 equality constraints.

If we denote $\alpha_2 = bm$, we can use the variable b instead of α_2.

The fact that the least favorable prior is an equalizer rule (13) shows that this problem is equivalent to the maximization of the convex combination of the several g_i function without the constraint on θ. The problem, we eventually studied is the following:

$$\begin{cases} \text{Minimize} : -\{a_1 g_1 + a_2 g_2 + a_3 g_3\}, \\ \text{Subject to} : \\ \qquad a_1 + a_2 + a_3 = 1, \\ \qquad 0 \le a_i \le 1, \quad i = 2, 3, \\ \qquad 0 \le b \le 1, \end{cases} \tag{16}$$

where:

$$g_1(a_2, a_3, b) : \delta(1)^2 = \theta,$$

$$g_2(a_2, a_3, b) : bm \exp(-bm) + \sum_{x=1}^{\infty} (bm - \delta(x))^2 \frac{(bm)^{x-1}}{x!} \exp(-bm) = \theta,$$

$$g_3(a_2, a_3, b) : m \exp(-m) + \sum_{x=1}^{\infty} (m - \delta(x))^2 \frac{m^{x-1}}{x!} \exp(-m) = \theta,$$

and

$$\begin{cases} \delta(x) = m \dfrac{a_2 b^x \exp(1-b)m + a_3}{a_2 b^{x-1} \exp(1-b)m + a_3}, \text{if } x \neq 0, 1, \\ \delta(1) = m \dfrac{a_2 b \exp(-bm) + a_3 \exp(-m)}{1 - a_2 + a_3 + a_2 \exp(-bm) + a_3 \exp(-m)}. \end{cases}$$

The algorithm of stochastic perturbation of reduced gradient (see, for instance, [1]) used to solve the statistical problem of estimating a bounded mean with a minimax procedure.

5. Random Perturbation of Reduced Gradient Method

5.1. Reduced gradient method

We consider the following problem:

$$\begin{cases} \text{Minimize } f(\mathbf{x}), \\ \text{subject to } A\mathbf{x} = b, \\ 0 \leq \mathbf{x}, \end{cases} \tag{17}$$

where f is twice continuously differentiable function, A is $m \times n$ matrix and b is a vector in $\mathbb{R}^m$. By assumption, the matrix A has full row rank. Let a feasible solution $\mathbf{x}^k \geq 0$, and let us assume that a basis B, where $\mathbf{x}_B^k > 0$. The reduced gradient method begins with a basis B and a feasible solution $\mathbf{x}^k = (\mathbf{x}_B^k, \mathbf{x}_N^k)$.

Now, let us assume that the basis is non degenerate, i.e., only the non negativity constraints $\mathbf{x}_N \geq 0$ might be active at the current iterate $\mathbf{x}^k$. Let the search direction be a vector $d = (d_B^t, d_N^t)^T$ in the null space of the matrix A, defined as $d_B = -B^{-1} N d_N$ and $d_N \geq 0$. If we define so, then the feasibility of $\mathbf{x}^k + \eta d$ is guaranteed as long as $\mathbf{x}_B^k + \eta d_B \geq 0$, i.e., as long as

$$\eta \leq \eta_{\max} = \min_{i \in B, d_i < 0} \{ \frac{-x_i^k}{d_i} \}. \tag{18}$$

We still need to define $d_N \geq 0$ such that it is a descent direction of f_N projected to the coordinate hyperplane active at the current point x_N^k.

So, we have

$$d_j = \begin{cases} 0, \text{ if } x_j^k = 0 \text{ and } \dfrac{\partial f_N(\mathbf{x}_N^k)}{\partial x_j} \geq 0, \\ -\dfrac{\partial f_N(\mathbf{x}_N^k)}{\partial x_j}, \text{otherwise,} \end{cases} \qquad j \in N,$$

where $f_N(\mathbf{x}_N) = f(\mathbf{x}) = f(B^{-1}b - B^{-1}N\mathbf{x}_N, \mathbf{x}_N)$.

To complete the description of the algorithm, we make a line search to obtain the new point.

$$\mathbf{x}^{k+1} = Q(\mathbf{x}^k) = \underset{0 \leq \eta \leq \eta_{\max}}{\arg\min} f(\mathbf{x}^k + \eta d_k). \tag{19}$$

If all the coordinates $\mathbf{x}_B^{k+1}$ stay strictly positive, we keep the basis, else a pivot is made to eliminate the zero variable from the basis and replace it by a positive, but currently non basic coordinate.

5.2. Stochastic perturbation

The main difficulty remains the lack of convexity, if f is not convex, the Kuhn-Tucker points may not correspond to global minima. In the sequel, we shall improve this point by using an appropriate random perturbation.

The sequence of real numbers $\{\mathbf{x}^k\}_{k \geq 0}$ is replaced by a sequence of random variables $\{X^k\}_{k \geq 0}$ involving a random perturbation $\mathcal{P}_k$ of the deterministic iteration (19). A simple strategy consists in

$$X^0 = \mathbf{x}^0; \ \forall k \geq 0 \ X^{k+1} = Q_k(X^k) + \mathcal{P}_k. \tag{20}$$

Equation (20) can be viewed as perturbation of the descent direction d_k, which is replaced by a new direction $D_k = d_k + \mathcal{P}_k / \eta_k$, and the iterations (19) become

$$X^{k+1} = X^k + \eta_k D_k.$$

The procedure generates a sequence $U_k = f(X^k)$. By construction, this sequence is decreasing and lower bounded by U^*, and converge to the global minimum (see, for instance, [1], [2], and [13]).

6. Numerical Result

By using random perturbation of reduced gradient method for solving global optimization problem (16) with initial point $(a_1 = 0.1, a_2 = 0.4, a_3 = 0.5, b = 0.5)$ and number of perturbation = 2000, we find the following results, see Table 1, where NI is the number of iterations.

Table 1. Least favorable distributions and minimax risk on $[0, m]$ as a function of m

m	a_1	a_2	a_3	b	α_2	θ	NI
2.7	0.09389	0.90702	0.0000	0.50784	1.37	–0.40498	8
3	0.09530	0.90568	0.0000	0.45531	1.37	–0.40488	6
3.5	0.09316	0.90722	0.0000	0.36173	1.27	–0.40425	7
4	0.09316	0.90760	0.0000	0.32486	1.30	–0.40492	6
4.5	0.09432	0.90652	0.0000	0.30125	1.36	–0.40521	5
5	0.09899	0.90182	0.0000	0.27243	1.36	–0.40505	6

References

[1] A. El Mouatasim, Reduced gradient method and its generalization via stochastic perturbation, Journal of Natural Science and Mathematics 2(1) (2008), 17-40.

[2] A. El Mouatasim, R. Ellaia and J. E. Souza de Cursi, Random perturbation of variable metric method for unconstraint nonsmooth nonconvex optimization, J. Appl. Math. Comp. Sci. 16(4) (2006), 463-474.

[3] M. N. Ghosh, Uniform approximation of minimax point estimates, Ann. Math. Statist. 35 (1964), 1031-1047.

[4] M. B. Gonalves and J. E. Souza de Cursi, Parameter estimation in a trip distribution model by random perturbation of a descent method, Transport Research Part B 35 (2001), 137-161.

[5] E. Gourdin, Global Optimization Algorithms for the Construction of Least Favorable Priors and Minimax Estimators, Engineering Undergraduate Thesis, Institut d'Informatique d'Interprise, Every, France, 1989.

[6] E. Gourdin, B. Jaumard and K. B. MacGibbon, Global optimization decomposition methods for bounded parameter minimax evaluation, SIAM J. Sci. Comput. 15(1) (1994), 16-35.

[7] H. Hara and A. Takemura, Bayes admissible estimation of the means in Poisson decomposable, Journal of Statistical Planning and Inference 139 (2009), 1297-1319.

[8] I. M. Johnstone and K. B. MacGibbon, Minimax estimation of a constrained Poisson vector, Ann. Stat. 20(2) (1992), 807-831.

[9] I. M. Johnstone and K. B. MacGibbon, Asymptotically minimax estimation of a constrained Poisson vector via polydisc transforms, Annales de l'I.H.P., Section B, Tome 29(2) (1993), 289-319.

[10] É. Marchand and A. Parsian, Minimax estimation of a bounded parameter of a discrete distribution, Stat. Prob. Lett. 76 (2006), 547-554.

[11] A. M. Mood, F. A. Graybill and D. C. Boes, Introduction to the Theory of Statistics, Third Edition, McGraw-Hill, Statistical Series, 1974.

[12] A. Papoulis and S. U. Pillai, Probability, Random Variables and Stochastic Processes, Tata McGraw-Hill Edition, 2002.

[13] P. Sadegh and J. C. Spall, Optimal random perturbations for stochastic approximation using a simultaneous perturbation gradient approximation, IEEE Trans. Autom. Control 44(1) (1999), 231-232.

■