EESTI NSV TEADUSTE AKADEEMIA TOIMETISED. 24. KÖIDE
FÜÜSIKA * MATEMAATIKA. 1975, NR. 1

ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ТОМ 24
ФИЗИКА * МАТЕМАТИКА. 1975, № 1

https://doi.org/10.3176/phys.math.1975.1.01

УДК 512.25/26 + 519.3 : 330.115

*Э. РАЙК*

# ДИФФЕРЕНЦИРУЕМОСТЬ ПО ПАРАМЕТРУ ФУНКЦИИ ВЕРОЯТНОСТИ И СТОХАСТИЧЕСКИЙ ПСЕВДОГРАДИЕНТНЫЙ МЕТОД ДЛЯ ЕЕ ОПТИМИЗАЦИИ

Пусть на $n$-мерном пространстве $X$ задана функция вероятности

$$v(x) = P\{f(x, \xi(\omega)) \geqslant 0\}, \tag{1}$$

где $\xi(\omega)$ — случайный вектор со значениями в $m$-мерном пространстве $Y$, $f(x, y)$ — действительная функция, определенная на произведении пространств $X \times Y$.

Допустим ниже, что случайный вектор $\xi(\omega)$ индуцирует в пространстве $Y$ вероятностное распределение, которое имеет лишь абсолютно непрерывное составляющее и, следовательно, полностью определяется своей функцией плотности вероятности $p(y)$. Это позволяет функцию вероятности выписать через $m$-кратный интеграл

$$v(x) = \int\limits_{f(x,y)\geqslant 0} \int \ldots \int p(y)\, dy. \tag{2}$$

При решении задач нелинейного стохастического программирования, содержащих функции вероятности в качестве функций ограничения или функции цели, встает вопрос о виде градиента и условий дифференцируемости функции вероятности.

Ответ на этот вопрос дает

Теорема 1. *Пусть* 1) *частные производные* $f'_x(x, y)$, $f'_y(x, y)$ *и плотность вероятности* $p(y)$ *непрерывны*;
2) *множество* $S(x, 0) = \{y : f(x, y) = 0\}$ *равномерно ограничено в некоторой окрестности точки* $x$;
3) *функция* $f'_y(x, y)$ *удовлетворяет условию Липшица по* $y$ *и функция* $\|f'_y(x, y)\|$ *положительна на множестве* $S(x, 0)$;
4) *поверхностный интеграл* $\int\limits_{S(x,0)} \int \ldots \int dS$ *существует.*

*Тогда функция вероятности* $v(x)$ *дифференцируема по Фреше в точке* $x$ *и градиент* $v'(x)$ *равняется* $m-1$*-кратному интегралу*

$$v'(x) = \int\limits_{S(x,0)} \int \ldots \int f'_x(x, y)\, p(y)\, \|f'_y(x, y)\|^{-1}\, dS. \tag{3}$$

Доказательство. В силу предположения 2 и непрерывности функции $f(x, y)$ поверхность $S(x + \Delta x, 0) = \{y : f(x + \Delta x, y) = 0\}$ приближается к поверхности $S(x, 0)$ равномерно в том смысле, что максимальное расстояние $\varrho(\Delta x)$ точек поверхности $S(x + \Delta x, 0)$ до поверхности $S(x, 0)$ стремится к нулю, т. е. $\varrho(\Delta x) \to 0$ при $\|\Delta x\| \to 0$. Действительно, так как множества $S(x, 0)$ и $S(x + \Delta x, 0)$ компакты, расстояние $\varrho(\Delta x)$ существует и равняется

$$\varrho(\Delta x) = \max_{u \in S(x+\Delta x, 0)} \min_{w \in S(x,0)} \|u - w\|.$$

Предположим обратное: $\varrho(\Delta x) \not\to 0$ при $\|\Delta x\| \to 0$. Тогда найдется такая сходящаяся последовательность точек $\{x_n, y_n\}$, реализующих расстояние

$$\varrho(x_n - x) = \min_{w \in S(x,0)} \|y_n - w\|, \qquad y_n \in S(x_n, 0),$$

что $f(x_n, y_n) = 0$ и по непрерывности $f(x, y) = 0$, но это противоречит предположению $\varrho(\Delta x) \not\to 0$.

При непрерывных частных производных $f'_x(x, y)$ и $f'_y(x, y)$ имеем разложение

$$f(x+\Delta x, y+\Delta y) = f(x, y) + (f'_x(\tilde{x}, \tilde{y}), \Delta x) + (f'_y(\tilde{x}, \tilde{y}), \Delta y), \tag{4}$$

где $$\tilde{x} = x + \Theta(y)\Delta x, \qquad \tilde{y} = y + \Theta(y)\Delta y, \qquad 0 \leqslant \Theta(y) \leqslant 1.$$

Здесь через $(x, y)$ обозначено скалярное произведение векторов $x$ и $y$. Функцию $\Theta(y)$ можно считать измеримой ([1], с. 409). Приравняем в формуле (4) $f(x, y) = 0$ и $f(x + \Delta x, y + \Delta y) = 0$ и выразим из формулы (4) приращение $\Delta y$, взятое параллельно вектору нормали поверхности $S(x, 0)$ в точке $y$, через

$$\Delta y = -\frac{f'_y(x, y)\,(f'_x(\tilde{x}, \tilde{y}), \Delta x)}{(f'_y(\tilde{x}, \tilde{y}), f'_y(x, y))}.$$

В силу условия 3 теоремы существует $\Delta x$ такая, начиная с которой $(f'_y(\tilde{x}, \tilde{y}), f'_y(x, y)) \geqslant \gamma > 0$ для всех $y$ поверхности $S(x, 0)$. Проекция приращения $\Delta y$ на внешнюю нормаль $-f'_y(x, y)\|f'_y(x, y)\|^{-1}$ множества $\{y : f(x, y) \geqslant 0\}$ равняется

$$l(y) = \frac{(f'_x(\tilde{x}, \tilde{y}), \Delta x)\,\|f'_y(x, y)\|}{(f'_y(\tilde{x}, \tilde{y}), f'_y(x, y))}$$

Функция $l(y)$ имеет вполне определенный геометрический смысл. Она равняется по абсолютной величине отрезку между поверхностями $S(x, 0)$ и $S(x + \Delta x, 0)$, взятому параллельно вектору нормали поверхности $S(x, 0)$. Функция $l(y)$ является положительной, если отрезок расположен вне множества $\{y : f(x, y) \geqslant 0\}$, и отрицательной, если отрезок расположен внутри множества.

Приращение функции вероятности $v(x)$ выражается в виде разности $m$-кратных интегралов от функции $p(y)$

$$v(x+\Delta x) - v(x) = \int\!\!\int \ldots \int_{V_1} p(y)\,dy - \int\!\!\int \ldots \int_{V_2} p(y)\,dy,$$

где
$$V_1 = \{y : f(x+\Delta x, y) \geqslant 0,\ f(x, y) < 0\}$$

и
$$V_2 = \{y : f(x+\Delta x, y) < 0,\ f(x, y) \geqslant 0\}.$$

Эти $m$-кратные интегралы можно выразить через $m-1$-кратный поверхностный интеграл

$$v(x+\Delta x) - v(x) = \int \ldots \int_{S(x,0)} l(y)\,p(h(y))\,dS + o(\|\Delta x\|),$$

где $$h(y) = y + f'_y(x, y)\|f'_y(x, y)\|^{-1} l(y)\,\Theta_1(y), \qquad 0 \leqslant \Theta_1(y) \leqslant 1.$$

Это выражение можно получить непосредственно, учитывая то, что объем элементарного кривого цилиндра $\Delta V$, боковая поверхность которого образуется нормалями поверхности $S(x, 0)$, с точностью до бесконечно малых высшего порядка относительно $\|\Delta x\|$, равняется $\Delta V = \Delta S_x l(\bar{y})$, $\bar{y} \in S_x$, где $\Delta S_x$ — площадь основания элементарного цилиндра.

По определению функция $l(y)$ положительна для области $V_1$ и отрицательна для области $V_2$.

Итак,

$$v(x+\Delta x)-v(x)=\int\limits_{S(x,0)}\dots\int(f'_x(x,y),\Delta x)\,p(y)\,\|f'_y(x,y)\|^{-1}dS+\omega(x,\Delta x),$$

причем остаточный член равняется

$$\omega(x,\Delta x)=\int\limits_{S(x,0)}\dots\int\frac{(f'_x(\tilde x,\tilde y),\Delta x)\,p(h(y))\,\|f'_y(x,y)\|}{(f'_y(\tilde x,\tilde y),f'_y(x,y))}dS-$$

$$-\int\limits_{S(x,0)}\dots\int\frac{(f'_x(x,y),\Delta x)\,p(y)}{\|f'_y(x,y)\|}dS+o(\|\Delta x\|),$$

$$\tilde x=x+\Theta(y)\Delta x,\qquad h(y)=y+\frac{f'_y(x,y)\,(f'_x(\tilde x,\tilde y),\Delta x)}{(f'_y(\tilde x,\tilde y),f'_y(x,y))}\Theta_1(y),$$

$$\tilde y=y+\Theta(y)\Delta y,$$

$$0\leqslant\Theta(y)\leqslant1,\qquad 0\leqslant\Theta_1(y)\leqslant1.$$

Покажем, что выбирая $\|\Delta x\|$ малой, отношение $\dfrac{|\omega(x,\Delta x)|}{\|\Delta x\|}$ можно сделать сколь угодно малым. Допустим, что $\|\Delta x\|$ настолько мала, что $(f'_y(\tilde x,\tilde y),\ f'_y(x,y))\geqslant\gamma>0$ на множестве $Q=\{y:\|z-y\|\leqslant\varrho(\Delta x),\ z\in S(x,0)\}$. Обозначим предельные значения непрерывных функций на компакте через

$$\max_{y\in Q}(\|f'_x(x,y)\|,\ \|f'_y(x,y)\|,\ p(y))=K_1,\ \min_{y\in Q}\|f'_y(x,y)\|=K_2.$$

Непрерывные функции $f'_x(x,y)$, $\|f'_y(x,y)\|$, $p(y)$ на компакте $Q$ равномерно непрерывны, следовательно, для любого $\varepsilon>0$ существует $\delta>0$ такое, что если выбрать $\|\Delta x\|$ настолько малой, что $\varrho(\Delta x)\leqslant\delta$, то

$\max\limits_{\|\tilde y-y\|\leqslant\|\Delta y\|}(\|f'_x(\tilde x,\tilde y)-f'_x(x,y)\|,|\|f'_y(\tilde x,\tilde y)\|-\|f'_y(x,y)\||,|p(h(y)-p(y)|)\leqslant\varepsilon.$

Тогда остаточный член оценивается таким образом

$$\frac{|\omega(x,\Delta x)|}{\|\Delta x\|}\leqslant\int\limits_{S(x,0)}\dots\int\left\|\frac{f'_x(\tilde x,\tilde y)\,p(h(y))}{\|f'_y(\tilde x,\tilde y)\|}-\frac{f'_x(x,y)\,p(y)}{\|f'_y(x,y)\|}\right\|dS\leqslant$$

$$\leqslant\int\limits_{S(x,0)}\dots\int dS\,\frac{3K_1^2}{K_2^2}\,\varepsilon.$$

По предположению теоремы поверхностный интеграл $\int\limits_{S(x,0)}\dots\int dS$ существует. Теорема доказана.

Наиболее жестким из условий теоремы 1 является требование ограниченности множества $S(x,0)=\{y:f(x,y)=0\}$. Если отказаться от этого требования, то верна

Теорема 2. *Пусть* 1) *частная производная* $f'_x(x,y)$ *и плотность вероятности* $p(y)$ *непрерывны*;

2) *функция* $f'_y(x,y)$ *удовлетворяет условию Липшица по* $y$ *и функция* $\|f'_y(x,y)\|$ *положительна на множестве* $S(x,0)$;

3) *существует поверхностный интеграл*

$$\int\limits_{S(x,0)}\dots\int f'_x(x,y)\,p(y)\,\|f'_y(x,y)\|^{-1}dS;$$

4) *для любого ограниченного множества* $Q$ *существует интеграл* $\int\limits_{S(x,0)\cap Q} \dots \int dS$;

5) *для любого* $\varepsilon > 0$ *существует ограниченное множество* $M \subset Y$ *и* $\delta > 0$ *такие, что если* $\|\Delta x\| \leqslant \delta$, *то* $|v_{Y\setminus M}(x+\Delta x) - v_{Y\setminus M}(x)| \leqslant \varepsilon\|\Delta x\|$, *где* $v_{Y\setminus M}(x) = \int\limits_{Y\setminus M\cap\{y: f(x,y)=0\}} \dots \int p(y)\,dy$.

*Тогда функция вероятности* $v(x)$ *дифференцируема по Фреше в точке* $x$ *и градиент* $v'(x)$ *равняется* $m-1$*-кратному интегралу*

$$v'(x) = \int\limits_{S(x,0)} \dots \int f'_x(x,y)\,p(y)\,\|f'_y(x,y)\|^{-1}\,dS.$$

Доказательство. Зафиксируем $\varepsilon > 0$. Тогда существует ограниченное множество $M$ и $\delta_1 > 0$ такие, что если $\|\Delta x\| \leqslant \delta_1$, то

$$|v_{Y\setminus M}(x+\Delta x) - v_{Y\setminus M}(x)| = |P\{f(x+\Delta x, \xi(\omega)) \geqslant 0,\ \xi(\omega) \in Y\setminus M\} - P\{f(x,\xi(\omega)) \geqslant 0,\ \xi(\omega) \in Y\setminus M\}| \leqslant \varepsilon\|\Delta x\|$$

и

$$\left|\int\limits_{S(x,0)\cap Y\setminus M} \dots \int f'_x(x,y)\,p(y)\,\|f'_y(x,y)\|^{-1}\,dS\right| \leqslant \varepsilon.$$

Имеем

$$\left|v_{Y\setminus M}(x+\Delta x) - v_{Y\setminus M}(x) - \int\limits_{S(x,0)\cap Y\setminus M} \dots \int (f'_x(x,y),\Delta x)\,p(y)\,\|f'_y(x,y)\|^{-1}\,dS\right| \leqslant 2\varepsilon\|\Delta x\|.$$

Согласно теореме 1, на множестве $M$ можем выбрать $\delta_2 > 0$ такое, что при $\|\Delta x\| \leqslant \delta_2$

$$\left|v_M(x+\Delta x) - v_M(x) - \int\limits_{S(x,0)\cap M} \dots \int (f'_x(x,y),\Delta x)\,p(y)\,\|f'_y(x,y)\|^{-1}\,dS\right| \leqslant \varepsilon\|\Delta x\|.$$

Итак, для любого $\varepsilon > 0$ существует $\delta = \min(\delta_1, \delta_2)$ такое, что

$$\left|v(x+\Delta x) - v(x) - \int\limits_{S(x,0)} \dots \int (f'_x(x,y),\Delta x)\,p(y)\,\|f'_y(x,y)\|^{-1}\,dS\right| \leqslant 3\varepsilon\|\Delta x\|.$$

Рассмотрим задачу минимизации функции вероятности

$$\min_{x\in X} v(x).$$

Использование градиентного метода для решения этой задачи затруднительно, так как приведенная формула (3) для вычисления градиента функции вероятности влечет за собой громоздкие вычисления. Помимо вычисления сложного интеграла по $m-1$-мерной поверхности, нам потребуется еще точно знать явный вид функции плотности вероятности $p(y)$. Оказывается, что эти трудности можно обойти.

Предлагаемый ниже стохастический псевдоградиентный метод для минимизации по параметру функции вероятности $v(x) = P[f(x,\xi(\omega)) \geqslant 0]$ не требует ни знания плотности вероятности $p(y)$, ни проведения громоздких вычислений. В этом смысле он напоминает известный градиентный алгоритм стохастической аппроксимации Роббинса—Монро [2].

Алгоритмы стохастической аппроксимации для решения задач, содержащих функции математического ожидания, нашли широкое использование начиная с работ [2, 3]; условия их сходимости хорошо изучены. В работе [4] исследована сходимость целого класса алгоритмов стохастической оптимизации, так наз. стохастических псевдоградиентных алгоритмов.

Приведем общий вид стохастических алгоритмов

$$x_k = x_{k-1} - \gamma_k s_k, \tag{5}$$

где $x_k \in X$, $\gamma_k$ — детерминированные скалярные множители, зависящие

только от $k$, $s_k$ — реализации случайного вектора в $X$. Следуя работе [4], назовем стохастический алгоритм для функции $v(x)$ псевдоградиентным, если

$$(v'(x_k),\ Es_k) \geqslant 0, \tag{6}$$

где $Es_k$ — математическое ожидание случайного вектора $s_k$.

Для функции вероятности $v(x)$ определим $s_k$ следующим образом:

$$s_k = f'_x(x_k, y_k), \tag{7}$$

где $y_k$ — первая очередная реализация случайного вектора, для которой

$$|f(x_k, y_k)| \leqslant \varepsilon_k, \qquad \varepsilon_k > 0. \tag{8}$$

Смысл и выбор значений $\varepsilon_k$ уточняется ниже.

Теорема 3. *Пусть выполнены условия теоремы* 1. *Тогда существует последовательность чисел* $\delta_k > 0$ *такая, что если* $\varepsilon_k \leqslant \delta_k$, *то алгоритм* (5), (7), (8) *является псевдоградиентным.*

Доказательство. Пусть $v'(x_k) \neq 0$. В противном случае неравенство (6) выполняется автоматически. Используя определение $s_k$, выразим математическое ожидание $Es_k$ через $m$-кратный интеграл

$$Es_k = \underset{|f(x_k,y)|\leqslant\varepsilon_k}{\int\int\ldots\int} f'_x(x_k, y)\, p_k(y)\, dy. \tag{9}$$

Здесь $p_k(y)$ определяется через $p(y)$ с учетом того, что используется лишь та реализация $y_k$, для которой выполняется неравенство (8):

$$p_k(y) = \frac{p(y)}{P(x_k, \varepsilon_k)} \quad \text{и} \quad p_k(y) \equiv 0, \ \text{если} \ P(x_k, \varepsilon_k) = 0,$$

где $P(x_k, \varepsilon_k) = \underset{|f(x_k,y)|\leqslant\varepsilon_k}{\int\int\ldots\int} p(y)\, dy.$

Используя теорему 108 [5], перепишем $m$-кратный интеграл (9) в виде

$$\underset{|f(x_k,y)|\leqslant\varepsilon_k}{\int\int\ldots\int} f'_x(x_k, y)\, p_k(y)\, dy = \int_{-\varepsilon_k}^{\varepsilon_k} dw \int_{S(x_k,w)}\ldots\int f'_x(x_k, y)\, p_k(y)\, \|f'_y(x_k, y)\|^{-1}\, dS,$$

где $S(x_k, w) = \{y : f(x_k, y) = w\}$.

По теореме о среднем значении имеем

$$Es_k = 2\varepsilon_k \int_{S(x_k,\bar{w})}\ldots\int f'_x(x_k, y)\, p_k(y)\, \|f'_y(x_k, y)\|^{-1}\, dS, \qquad -\varepsilon_k \leqslant \bar{w} \leqslant \varepsilon_k$$

и аналогично получаем

$$P(x_k, \varepsilon_k) = 2\varepsilon_k \int_{S(x_k,\tilde{w})}\ldots\int p(y)\, \|f'_y(x_k, y)\|^{-1}\, dS, \qquad -\varepsilon_k \leqslant \tilde{w} \leqslant \varepsilon_k.$$

Так как непрерывные функции $f'_x(x_k, y)$, $\|f'_y(x_k, y)\|$, $p_k(y)$ на компакте равномерно непрерывны и поверхности $S(x_k, \bar{w})$ и $S(x_k, \tilde{w})$ приближаются к поверхности $S(x_k, 0)$ равномерно, то

$$Es_k = 2\varepsilon_k \int_{S(x_k,0)}\ldots\int f'_x(x_k, y)\, p_k(y)\, \|f'_y(x_k, y)\|^{-1}\, dS + o_1(\varepsilon_k),$$

где $o_1(\varepsilon_k)$ — бесконечно малая величина высшего порядка относительно $\varepsilon_k$ и

$$P(x_k, \varepsilon_k) = 2\varepsilon_k P_k + o_2(\varepsilon_k),$$

$$P_k=\int\limits_{S(x_k,0)}\ldots\int p(y)\,\|f'_y(x_k,y)\|^{-1}\,dS.$$

Заметим, что $P_k>0$, так как $\|f'_y(x_k,y)\|>0$ и по предположению, сделанному вначале, $v'(x_k)\neq 0$.

Перепишем $Es_k$ в форме

$$Es_k=\frac{1}{P_k}\int\limits_{S(x_k,0)}\ldots\int f'_x(x_k,y)\,p(y)\,\|f'_y(x_k,y)\|^{-1}\,dS+o_3(\varepsilon_k)=\frac{1}{P_k}\,v'(x_k)+o_3(\varepsilon_k).$$

Тогда выражение в (6) примет вид

$$(v'(x_k),Es_k)=\frac{1}{P_k}\,\|v'(x_k)\|^2+o(\varepsilon_k).$$

Следовательно, существует $\delta_k>0$ такое, что если $\varepsilon_k<\delta_k$, то $(v'(x_k),\ Es_k)\geqslant 0$. Теорема доказана.

Покажем, что значения $\delta_k$, при которых алгоритм (5), (7), (8) является псевдоградиентным, можно вывести и явно. Одна, хотя и довольно грубая оценка, определяется следующим образом.

Теорема 4. *Пусть существуют предельные значения*

$$M=\max_y\,(p(y),\ \|f'_y(x,y)\|,\ \|f'_x(x,y)\|,\ \|f''_{yy}(x,y)\|),$$

$$K=\min_y\,\|f'_y(x,y)\|,\qquad L=\max(L_1,\ L_2,\ L_3),$$

*где $L_1, L_2, L_3$ постоянные Липшица по $y$ функций $p(y)$, $f'_y(x,y)$ и $f'_x(x,y)$ соответственно и пусть*

$$T_k=\int\limits_{S(x_k,0)}\ldots\int dS\Big(\int\limits_{S(x_k,0)}\ldots\int p(y)\,dS\Big)^{-1}.$$

*Тогда алгоритм (5), (7), (8) будет псевдоградиентным, если*

$$\varepsilon_k\leqslant\delta_k=\min\left(1,\ \frac{K}{L_2},\ \frac{1}{C_k}\,\|v'(x_k)\|\right),\tag{10}$$

*где*

$$C_k=3T_kK^{-2}LC(2T_kKMLC+1),\tag{11}$$

$$C=m!\,(1+12K^{-4}M^3)+1.\tag{12}$$

Доказательство этой теоремы весьма громоздко и малоинтересно и поэтому здесь не приводится.

С вычислительной точки зрения предпочтительнее выбрать $\varepsilon_k$ как можно бóльшей. Чем больше $\varepsilon_k$, тем меньше реализаций $y_k$, не удовлетворяющих неравенству (8), потребуется отвергнуть. При этом уменьшится соответственно и объем вычислений. С другой стороны, из неравенства (10) следует, что для установления алгоритма (5), (7), (8) сходящимся псевдоградиентным алгоритмом необходимо выбрать $\varepsilon_k\to 0$ при $k\to\infty$.

Отметим, наконец, что требование $\delta_k\leqslant 1$ из условия (10) несущественно. Это требование введено лишь для упрощения оценки (10)—(12).

Автор выражает признательность И. Петерсену и Т. Тобиасу за ценные замечания.

## ЛИТЕРАТУРА

1. Красносельский М. А., Забрейко П. П., Пустыльник Е. И., Соболевский П. Е., Интегральные операторы в пространствах суммируемых функций, М., 1966.

2. Robbins H., Monro S., A stochastic approximation method, Annals Math. Stat., 22, No. 1, 400 (1955).
3. Kiefer E., Wolfowitz J., Stochastic estimation of the maximum of a regression function, Annals Math. Stat., 23, No. 3, 462 (1952).
4. Поляк Б. Т., Цыпкин Я. З., Псевдоградиентные алгоритмы адаптации и обучения, Автомат. и телемех., № 3, 45 (1973).
5. Шварц Л., Анализ, т. I, М., 1972.

*Институт кибернетики*
*Академии наук Эстонской ССР*

Поступила в редакцию
13/V 1974

*E. RAIK*

## TÕENÄOSUSFUNKTSIOONI DIFERENTSEERUVUS PARAMEETRI JÄRGI JA TEMA OPTIMEERIMINE STOHHASTILISE PSEUDOGRADIENDIMEETODI ABIL

Esitatakse tõenäosusfunktsiooni diferentseerimise tingimused parameetri järgi ja gradiendi kuju ning konstrueeritakse stohhastiline pseudogradiendimeetod tõenäosusfunktsiooni minimeerimiseks.

*E. RAIK*

## THE DIFFERENTIABILITY IN THE PARAMETER OF THE PROBABILITY FUNCTION AND OPTIMIZATION OF THE PROBABILITY FUNCTION VIA THE STOCHASTIC PSEUDOGRADIENT METHOD

The differentiability conditions of the probability function are given, the form of the gradient of this function is deduced and the stochastic pseudogradient method for optimisation presented.