DOI: https://doi.org/10.36910/6775-2524-0560-2021-43-10
УДК 51 – 005.31:519.8 + 004.04
Мамчич Тетяна Іванівна, к. ф.-м. н., доцент,
https://orcid.org/0000-0002-1934-1484
Ханін Олександр Григоррвич, к. ф.-м. н., доцент,
https://orcid.org/0000-0003-3472-9871
Мамчич І. Я., студент
Волинський національний університет імені Лесі Українки, м. Луцьк, Україна

РОЗВ'ЯЗОК ЗАДАЧ ОПТИМІЗАЦІЇ ЗАСОБАМИ ПРОГРАМИ R З ПРИКЛАДОМ ОЦІНКИ ЙМОВІРНІСНОГО РОЗПОДІЛУ

Мамчич Т. І., Ханін О. Г., Мамчич І. Я. Розв'язок задач оптимізації засобами програми R з прикладом оцінки ймовірнісного розподілу. В роботі розглядаються можливості програми для задач оптимізації. Ця програма разом з асоційованими пакетами використана для виконання алгоритмів лінійної, квадратичної та нелінійної оптимізації. Головна увага зосереджена на обчисленнях для модифікованого методу оцінки розподілу ймовірностей. Пропонується ефективна обчислювальна технологія, придатна для широкого застосування.
Ключові слова: програма R, оптимізація, модифікований χ^2 метод, розподіл ймовірностей.

Мамчич Т. И., Ханин О. Г., Мамчич И. Я. Решение задач оптимизации средствами программы R с примером оценки вероятностного распределения. В роботе рассмотрены возможности программы для задач оптимизации. Эта программа вместе с ассоциированными пакетами использована для выполнения алгоритмов линейной, квадратичной и нелинейной оптимізации. Головное внимание сосредоточено на вычислениях для модифицированного метода оценки распределения вероятностей. Предлагается эффективная вычислительная технология для широкого применения.
Ключовые слова: программа R , оптимизация, модифицированний χ^2 метод, распределение вероятностей.

Mamchych T.I., Khanin O.G., Mamchych I.Ya. Solution of Optimization Problems using program R together with the example estimation of probabilities distribution. Possibilities of the program R for problems of optimization are considered in this paper. The program together with associated packages is used for performing algorithms of linear, quadratic and nonlinear optimization. The main attention is focused on calculating for modified χ^2 method for estimation of probabilities distribution. Here we introduce an effective computational technology available for wide practical application.
Keywords:: program R, optimization, modified χ^2 method, probabilities distribution.

Постановка проблеми. Методи оптимізації наявні у вигляді самостійних дисциплін або у складі більш загального курсу в освіті не тільки математиків. Для інформатиків, інженерів, фахівців економічного профілю, бізнес-аналітиків методи знаходження найкращих у певному сенсі рішень відносяться до пріоритетних завдань, що і відображається у присутності таких тем у навчальних дисциплінах. Вивчаються як аналітичні методи знаходження оптимумів, так і чисельні. Для виконання обчислень залучаються наявні комп'ютерні програми. Найчастіше це MS Excel або MatLab. Названі програми тільки частково задовільняють потреби навчальних та дослідницьких задач, оскільки в Excel обчислення надто громіздкі, а програма MatLab у більшості випадків відсутня. Тому є потреба в доступних обчислювальних інструментах.

Мета статті. Дослідження має методологічний характер і полягає у вивченні можливостей програми R для розробки ефективного обчислювального інструменту аналізу емпіричних даних.

Аналіз наукових досліджень і публікацій. Ця робота виконана в межах ширшого завдання, яке полягало у розробці ефективного обчислювального алгоритму для знаходження оцінки теоретичного розподілу ймовірностей для реалізації модифікованого методу Пірсона, запропонованого в попередніх роботах Ханіна О.Г. [1], [2]. В цих же роботах наводиться й приклад обчислення реальних даних за допомогою програми MS Excel. Обчислення в MS Excel є досить громіздкими, а запропонований математичний метод для широкого вжитку потребує зручних обчислювальних інструментів. З цією метою було вивчено можливості програми R.

Виклад основного матеріалу й обґрунтування результатів дослідження. В даній роботі пропонується використовувати програму R з асоційованими пакетами “lpSolver”, ”quadprog” та ”nloptr”. Як програма R, так і вказані пакети є безкоштовними. Наведено приклади розв'язку задач лінійного програмування (LP), квадратичного програмування (QP) та задач нелінійного програмування.

Задачу LP виконано за допомогою команди "lp" (lpSolver), в якій параметри цільової функції, обмежень-рівностей, обмежень-нерівностей задано окремими опціями. Координати оптимального значення отримуються за допомогою lp$solution.

Приклад 1. Знайти розв'язок задачі лінійного програмування: Max F = x + 2 y

3 x - 2 y >= - 4;
y <= 5;
x + 2 y <= 14;
2 x - y <= 8.

Виконання: код і результат виконання.

```
> library("lpSolve")
> f.obj= c(1,2)
> f.con=matrix(c(3, -2, 0, 1, 1, 2, 2, -1), nrow=4, byrow=TRUE)
> f.dir=c(">=","<=","<=")
> f.rhs=c(-4, 5, 14, 8)
> lp("max",f.obj,f.con,f.dir,f.rhs)
Success: the objective function is 14
> lp("max",f.obj,f.con,f.dir,f.rhs)$solution
[1] 6 4
```

Отже, $f_{max}(6,4) = 14$.

В задачі QP використано команду "solve.QP" (quadprog) з параметрами, де цільова функція має стандартне представлення у вигляді суми лінійної та квадратичної частини, а також лінійні обмеження представлено вектором коефіцієнтів.

Приклад 2. Знайти розв'язок задачі квадратичного програмування:

Min $F = x^2 + y^2 - 2x - 2y$ s.t. $2x - y >= -1$; $x - 3y >= -8$; $x + y <= 8$; $2x - y <= 10$.

Функція обчислює розв'язок задачі Min $(- d^T b + ½ b^T D b)$ при умові $A^T b >= b0$.

Тому перед її застосуванням потрібно подати систему обмежень у формі

$2x - y >= -1$; $x - 3y >= -8$; $- x - y >= - 8$; $- 2x + y >= -10$.

Виконання: код і результат виконання.

```
> library("quadprog")
> Dmat=matrix(c(2,0,0,2),2,2)
> dvec=c(2,2)
> Amat=matrix(c(2,-1,1,-3,-1,-1,-2,1),2,4)
> bvec=c(-1, -8, -8, -10)
> solve.QP(Dmat, dvec, Amat, bvec=bvec)
$solution
[1] 1 1
$value
[1] -2
$unconstrained.solution
[1] 1 1
```

Отже, $f_{min}(1,1) = -2$.

В наведеному прикладі досліджуються задачі з галузі маркетингу. Кожен об'єкт представлено емпіричним розподілом частот. Такого типу дані досить типові при вивченні преференцій респондентів, коли робиться вибір серед запропонованих альтернатив. В модифікованому методі Пірсона ([1]) пропонується для кожного об'єкта знайти оцінку теоретичного розподілу ймовірностей як розв'язок задачі мінімізації, де цільовою функцією є величина хі-квадрат - відхилення емпіричного розподілу частот від деякого заданого (еталонного) розподілу, а система обмежень складається з умови p1 + p2 + ... + pm = 1 та умови потрапляння pi у відповідні довірчі інтервали, обчислені за емпіричними частотами для заданої надійності.

Для цієї задачі було використано пакет "nloptr". Ця функція обчислює локальний мінімум градієнтним чисельним методом. Алгоритм розрахунків для контролю виконано для тих же даних, що і в публікації [1]. Отримано результат, близький до наведеного попередньо, і навіть точніший.

Приклад 3.

Розмір еталонної вибірки: n = 283.

Емпіричні частоти: vi = {34,103,24,90,14,18}.

Довірчі інтервали для p1, p2, ... , p6 :

(0.245; 0.313), (0.223;0.288), (0.084;0.130), (0.156;0.214),

(0.009; 0.029), (0.127; 0.181).

Задача полягає в мінімізації цільової функції:

X^2 (p1, p2, ... , p6) = ((34 - n*p1)^2/p1 + ((103 - n*p2)^2/p2 + ((24 - n*p3)^2/p3 + ((90 - n*p4)^2/p4 + ((14 - n*p5)^2/p5 + ((18 - n*p6)^2/p6)/n

при заданих обмеженнях:

p1 + p2 + p3 + p4 + p5 + p6 = 1; 0.245 <= p1 <= 0.313; 0.223 <= p2 <= 0.288;

0.084 <= p3 <= 0.130; 0.156 <= p4 <= 0.214; 0.009 <= p5 <= 0.029; 0.127 <= p6 <= 0.181.

Виконання: код скрипту і результат виконання:

```
n=681
v=c(34,103,24,90,14,18)
lb=c(0.245, 0.223, 0.084, 0.156, 0.009, 0.127)
ub=c(0.313,0.288,0.130,0.214,0.029,0.181)
eval_f=function(x){
    return( list("objective" = (1/n)*((v[1] - n*x[1])^2/x[1] + (v[2] - n*x[2])^2/x[2] +(v[3] - n*x[3])^2/x[3] + (v[4] - n*x[4])^2/x[4] + (v[5] - n*x[5])^2/x[5] + (v[6] - n*x[6])^2/x[6]),
        "gradient" = c(n - v[1]*v[1]/(n*x[1]*x[1]),
                    n - v[2]*v[2]/(n*x[2]*x[2]),
                    n - v[3]*v[3]/(n*x[3]*x[3]),
                    n - v[4]*v[4]/(n*x[4]*x[4]),
                    n - v[5]*v[5]/(n*x[5]*x[5]),
                    n - v[6]*v[6]/(n*x[6]*x[6]))))
}
eval_g_eq = function(x){
        constr = c(x[1]+x[2]+x[3]+x[4]+x[5]+x[6]-1)
        grad = c(1,1,1,1,1,1)
        return(list("constraints"=constr, "jacobian"=grad))
}
x0 = lb
local_opts = list("algorithm" = "NLOPT_LD_MMA",
            "xtol_rel" = 1.0e-7)
opts = list("algorithm" = "NLOPT_LD_AUGLAG",
        "xtol_rel" = 1.0e-7,
        "maxeval" = 1000,
        "local_opts" = local_opts)
res = nloptr(x0=x0,
        eval_f=eval_f,
        lb=lb,
        ub=ub,
        eval_g_eq=eval_g_eq,
        opts=opts)
print(res)
Optimal value of objective function:  253.992049817991
Optimal value of controls: 0.245 0.288 0.097 0.214 0.029 0.127
```

Висновки та перспективи подальших досліджень.

Можливості розглянутих пакетів значно ширші, ніж це проілюстровано в наведених прикладах. “Nloptr” містить команди для реалізації методів як градієнтних, так і не пов'язаних з похідними, зокрема, методу множників Лагранжа (емуляція), оптимізації лінійними апроксимаціями, стохастичних методів пошуку, методів пошуку глобальних оптимумів, а також цілий набір допоміжних процедур ([4]).

З огляду на отримані результати, можна визнати, що програма R разом із асоційованими пакетами “lpSolve” та “nloptr” є ефективним інструментом вирішення задач оптимізації в реальних дослідженнях. Тому вважаємо доцільним використання цієї програми в навчанні другого (магістерського) та третього освітнього рівнів для фахівців різних спрямувань, а також широке її використання у практиці наукових досліджень. Особливо, зважаючи на доступність для використання цієї безкоштовної програми високого якісного рівня.

Запропонований метод для знаходження розподілу ймовірностей разом із розробленою технологією обчислень в R може використовуватись в аналізі даних. Розподіл частот отримують в дослідженнях у різних галузях – політології, соціології, маркетингу – всюди, де вивчають преференції респондентів щодо вибору серед заданих альтернатив. Оцінка розподілу ймовірностей є наступним логічним кроком.

Варто зазначити також метод контент-аналізу (Text Mining), де частоти є базовими даними, для якого викладені технології також можуть застосовуватись.

В сучасних дослідженнях, пов'язаних із класифікаціями, у всіляких алгоритмах штучного інтелекту деталі чисельних процедур є вагомим фактором застосовності. Тому можна вважати, що дана розробка є корисною, оскільки вона уможливлює практичне застосування методу.

У нашому дослідженні класифікація отриманих ймовірнісних розподілів є наступним логічним кроком. Розглядаючи знайдені розподіли як презентацію кожного об'єкта, можна встановлювати групи "подібних" методами кластерного аналізу, виявляти приховані фактори впливуабо ж розв'язувати інші задачі багатовимірного аналізу.

Список бібліографічного опису

1. Майборода Р.Є. (2019). Комп'ютерна статистика: Навчальний посібник. – К.:ВПЦ «Київський університет»
2. Ханін О.Г.(2015). Методологічні особливості застосування критерію узгодженості Х2 в практичних задачах соціології, економіки та маркетингу. Економічний аналіз: зб. наук. праць Тернопільський національний економічний університет. Тернопіль. Видавничо-поліграфічний центр Тернопільського національного економічного університету "Економічна думка". 22, 1, 67 – 70.
3. Ханін О.Г. (2016). Метод Х2 - кластеризації в задачах маркетингу. Економічний аналіз: зб. наук. праць Тернопільський національний економічний університет. Тернопіль. Видавничо-поліграфічний центр Тернопільського національного економічного університету "Економічна думка", 26, 38 – 42.
4. Berkelaar M. Package 'lpSolver', https://cran.r-project.org/web/packages/lpSolve/lpSolve.pdf [Accessed: May 18, 2021].
5. Turlach B., Weingessel A. Package 'quadprog', https://cran.r-project.org/web/packages/quadprog/quadprog.pdf [Accessed: May 18, 2021].
6. Johnson S.G. The NLopt nonlinear-optimization package.
7. https://cran.r-project.org/web/packages/nloptr/citation.html [Accessed: April 1, 2021]
8. K. Pearson. (1900). On the criterion that a given system of deviations from the probable in case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine. 50, 157.

References

1. Maiboroda, R. E. (2016). Computer statistics. Textbook. Kyiv Taras Shevchenko National University
2. Khanin, O.G. (2015). Methodological features of X2 consistency criteria in the practical problems of Sociology, Economics and Marketing. Economic analysis, 22, 1, 67 – 70.
3. Khanin,O.G. (2016). The method of X2 clustering in problems of Marketing. Economic analysis, 26, 1, 38 – 42.
4. Berkelaar M. Package 'lpSolver', https://cran.r-project.org/web/packages/lpSolve/lpSolve.pdf [Accessed: May 18, 2021].
5. Turlach B., Weingessel A. Package 'quadprog', https://cran.r-project.org/web/packages/quadprog/quadprog.pdf [Accessed: May 18, 2021].
6. Johnson S.G. The NLopt nonlinear-optimization package, https://cran.r-project.org/web/packages/nloptr/citation.html [Accessed: April 1, 2021].
7. K. Pearson. On the criterion that a given system of deviations from the probable in case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine, 1900. 50, 157.