Team Control Number

10751

Problem Chosen

B

2020
HiMCM
Summary Sheet

Summary summary.

Contents

1   Introduction

The decrease in biodiversity due to the extinction of endangered plants and animals is a serious problem. To avoid this, people need to spend money on protecting the plants and animals.

However, in some places like Florida, greatly needing biodiversity conservation, people do not have enough money to protect all of the imperiled species there. They need to find out a proper scheme of funding and protecting to make full use of the funds to get as much benefit as possible.

Different projects of conservation have different cost, benefit, and required time, as shown in the table of threatened plants data (abbreviated as TPD in the following parts of the paper). It is to be determined which projects are selected in the optimal plan according to these factors. After choosing the projects, when to start the projects should also be determined to balance the funds spent over time as possible. To make it clear what is going to be done, the problem is restated mathematically in Section 3. The model and the method to derive the results are to be explained in Section 5.

We will write a non-technical memo in Section 8 to give our proposal according to our model.

2   List of symbols

To make our model concise and straight forward, a list of symbols are defined as following, as shown in Table 1. The meaning of the symbols are to be explained in detail in the rest of this article. In the table, the symbols in lower case are variables w.r.t. a project, and the symbols in upper case are variables w.r.t. the plan. The difference between a project and the plan is described in Section 5.1.

Note that in this article, all numberings (like x here numbering the projects, and n later numbering the years) start with 0 instead of 1 to meet the convention in computer science.

The symbol $\sum_j^n$ means to sum for $j = 0, 1, \ldots, n-1$.

3   Restatement of the problem

In Florida, many plant species are facing the challenge of extinction. Various plant conservation projects with different budgets and timelines are now being conducted to resolve this problem. However, one major issue is that funds mainly from the annual fundraising are not sufficient to support all projects. Therefore, it is essential to make a plan for the managers who monitor the whole plan to help them precisely decide which project should be included and when and how to allocate their fund. What's notable is that the final goal of this plan is not to simply save as many as plant species as possible. Instead, it should generate greatest benefit meanwhile taking account of different factors,

---

[1]The unit osu is invented to represent the unit of benefit.

Table 1: List of symbols

Symbol	Domain	Unit	Meaning
x	$\mathbb{N}$		The index of projects
D	$\mathbb{N}$	year	Duration of the plan
d_x	$\mathbb{N}$	year	Duration of project x
n	$\mathbb{N}$	year	Time (index of years)
Z	$\mathscr{P}(\mathbb{N})$		The set of chosen projects
F	$[0, +\infty)$	dollar	Total funds
b_x	$[0, +\infty)$	osu[1]	Benefit of project x
u_x	$[0, 1]$		Taxonomic uniqueness of project x
s_x	$[0, 1]$		Feasibility of success of project x
$c_{x,n}$	$[0, +\infty)$	dollar	Cost of project x in the nth year
c_x	$[0, +\infty)$	dollar/year	Total cost of project x
C_n	$[0, +\infty)$	dollar/year	Cost of the plan in the nth year
C	$[0, +\infty)$	dollar	Cost of the plan
β_x	$[0, +\infty)$	osu	Effective benefit of project x
B	$[0, +\infty)$	osu	Total effective benefit of the plan
T	$[0, +\infty)$	osu/year	Time efficiency of the plan
ξ	$\mathbb{N}$		The longest project in the plan
a_x	$\mathbb{N}$	year	When project x starts

like feasibility of success and the uniqueness of this species. The time spanning to finish the whole plan is also a factor that needs to be seriously considered because as more time being taken to protect one species, the likelihood of conserving another shall decrease, finally leading to less benefit. Another factor is the funds spent on these projects. Obviously, a project with median benefit but high expense doesn't fit. The most suitable plan we need to find out should include all variables mentioned and give the best result.

The mathematical description of the problem is given in Section 5. To be specific, the problem is to reach the goal described in Section 5.8 under the constraints described in Section 5.7.

4 General assumptions

Assumption 1. *The funds are provided off at one time.*

Assumption 2. *Conservation cannot be paused.*

Assumption 3. *Different conservations can be done simultaneously and do not affect each other.*

Assumption 4. *Extra funds' last year can be saved for this year.*

Assumption 5. *The conservation plan should not change when it is ongoing.*

Assumption 6. *The maximization of time efficiency is prior to the balancing of funds spent over time.*

5 The model

5.1 Projects and the plan

The word **project** is used to represent a conservation of a certain species of plant. According to the TPD, there are some **raw data** attributed to a project x, which are the **unique id**, the **taxonomic uniqueness** u_x, the **feasibility of success** s_x, the **cost** $c_{x,n}$ of the nth year. These concepts are explained in Section 5.2.

The word **plan** is used to represent how the projects are going to be executed, including which projects are executed (the set Z) and when they should start (the sequence $\{a_x\}$). Deciding the which projects to be executed is to maximize the time efficiency (described in Section 5.5). Deciding when the projects start is to minimize the variance of yearly cost (described in Section 5.6).

5.2 Raw attributes of projects

To make the concepts more clear and make it easy to use them later, although the meanings of the raw data are described in the TPD, their meanings should be re-described in a mathematical way.

Definition 1 (benefit)**.** *The **benefit** of project x is a real number $b_x \in [0, +\infty)$.*

The benefit is an important indicator indicating how imperiled the species is and how easy the project can be done. However, b_x cannot be directly used to measure how much can we benefit from finishing the project. Another derived concept called **effective benefit** should be used instead, as explained in Section 5.3.

Definition 2 (taxonomic uniqueness)**.** *The **taxonomic uniqueness** of project x is a real number $u_x \in [0, 1]$.*

The taxonomic uniqueness is a measure of how the species is unique from other species. It is stipulated that $u_x = 0$ if there exists a species that is exactly the same as the species conserved in project x, and $u_x = 1$ if the species conserved in project x does not have any similarities with any other species.

Definition 3 (feasibility of success)**.** *The **feasibility of success** of project x is a real number $s_x \in [0, 1]$.*

The meaning of s_x is the probability of the success of project x.

Definition 4 (yearly cost)**.** *The **yearly cost** of project x is a real number $c_{x,n} \in [0, +\infty)$.*

The meaning of $c_{x,n}$ is cost of project x in the nth year.

5.3 Effective benefit

Although b_x is defined for a project, hardly can a project lead to so much benefit finally. A new concept should be defined to describe how much can people actually benefit from finishing the project.

Definition 5 (effective benefit). *The **effective benefit** of a project is a real number $\beta_x \in [0, +\infty)$ defined as*

$$\beta_x := b_x u_x s_x. \tag{1}$$

Here is the explanation of why the effective benefit should be defined as Equation 1.

First, consider the effect of u_x. If $u_x = 0$, which means there is a species the same as the conserved species, the conservation is useless because even if the conserved species die out, there are other species taking the place. If $u_x = 1$, which means there are no similar species to the conserved species, the conservation is meaningful and can exert full benefit. From these two cases, it is reasonable to multiply b_x with u_x, which means u_x represents the portion of b_x that can be exerted if the project is successfully finished.

Then, consider the effect of s_x. It is a probability, so the effect should be considered statistically. Imagine there are many parallel universes, in each of which project x is carried out under the same condition. According to the law of large numbers, the portion of successful executions is the probability s_x, exerting the benefit, while the rest does not make any benefit. Then, the mean of all the benefits is $s_x \cdot b_x + (1 - s_x) \cdot 0 = b_x s_x$.

Combining the two effects above, the formula for effective benefit (Equation 1) can be derived.

From now on, when measuring how much can a project benefit people, β_x is used instead of b_x.

By summing up the effective benefit for all projects of the plan, the total effective benefit can be defined.

Definition 6 (total effective benefit). *The **total effective benefit** of the project is a real number $B \in [0, +\infty)$ defined as*

$$B := \sum_{x \in Z} b_x. \tag{2}$$

The B is an important indicator of how much a plan is expected to benefit if the plan is executed.

5.4 Duration

The **duration** of a project is the time for which it lasts. It can be defined according to the yearly costs of the project.

Definition 7 (duration of project). *The **duration of project** x is a real number $d_x \in \mathbb{N}$ defined as*

$$d_x := \min \left\{ m \in \mathbb{N} \middle| \forall n \geq m : c_{x,n} = 0 \right\} \tag{3}$$

According to the TPD, a project must be able to end, which means such d_x must exist.

The duration of the plan should be defined as the time interval between the time when the first project of the plan starts and the time when the last project of the plan ends. However, due to Assumption 6, the duration of the plan should unnecessarily be defined in a so straight-forward way. A better definition is given in Equation 6 in Section 5.5.

5.5 Time efficiency

Definition 8 (time efficiency)**.** *The **time efficiency** of the plan is a real number $T \in [0, +\infty)$ defined as*

$$T := \frac{B}{D}. \tag{4}$$

The time efficiency of the plan is the average benefit gain during a unit of time. It should be maximized to make the plan as efficient as possible, whose reason can be found in Section 5.8.

According to Assumption 6, T should be maximized before we think about how a_x can be adjusted to balance the funds spent over time. To maximize the time efficiency, the duration of the plan is the same as the duration of the project with the largest duration (ξ) in the plan, where ξ is the project with largest duration among Z

$$\xi := \arg\max_{x \in Z} d_x. \tag{5}$$

Then, when a_x are adjusted to minimize variance of yearly costs, the start time of projects with shorter duration than d_ξ will never be changed to make them start before ξ starts or end after ξ ends because otherwise D increases and thus makes T smaller. Therefore, rather than the straight-forward definition shown in Section 5.4, there is a better definition.

Definition 9 (duration of the plan)**.** *The **duration of the plan** is a real number $D \in \mathbb{N}$ defined as*

$$D := \max_{x \in Z} d_x. \tag{6}$$

It is obvious that, if C stays the same, the larger T is, the better the plan is. In fact, maximizing T is of the most priority. It may be surprising that the most prior goal is not maximizing $\frac{T}{C}$, which is more intuitive. We do not try to maximize $\frac{T}{C}$ because if we did that, it would lead to a trivial result: the plan only contains one project with the minimum $\frac{\epsilon_x}{d_x c_x}$, which is not what we want to see.

Proof. Without loss of generality, assume that there are only 2 plant candidates for the plan, and then the mathematical induction can be used to generalize the conclusion to any number of plant candidates.

Without loss of generality, assume that $d_0 \le d_1$, which means $D = d_1 \ge d_0$.

Case 1: $\frac{\varepsilon_0}{c_0} \leq \frac{\varepsilon_1}{c_1}$.

In this case,

$$\frac{\varepsilon_0}{c_0} \leq \frac{\varepsilon_0 + \varepsilon_1}{c_0 + c_1} \leq \frac{\varepsilon_1}{c_1}.$$

Thus, if the plan consists of both the 2 projects,

$$\frac{T}{C} = \frac{\varepsilon_0 + \varepsilon_1}{c_0 + c_1} \cdot \frac{1}{d_1} \leq \frac{\varepsilon_1}{c_1 d_1}.$$

Therefore, the plan consisting of only project 1 is better than that of both projects.

Case 2: $\frac{\varepsilon_0}{c_0} > \frac{\varepsilon_1}{c_1}$.

Similarly to case 1, in this case,

$$\frac{\varepsilon_0}{c_0} > \frac{\varepsilon_0 + \varepsilon_1}{c_0 + c_1} > \frac{\varepsilon_1}{c_1}.$$

Thus, if the plan consists of both the 2 projects,

$$\frac{T}{C} = \frac{\varepsilon_0 + \varepsilon_1}{c_0 + c_1} \cdot \frac{1}{d_1} < \frac{\varepsilon_0}{c_0 d_1} \leq \frac{\varepsilon_0}{c_0 d_0}.$$

Therefore, the plan consisting of only project 0 is better than that of both projects.

From the 2 cases above, it can be derived that there always exists a plan with single project having the largest $\frac{T}{C}$ compared to any other plan. $\square$

5.6 Variance of yearly costs

To measure how balanced the spent funds are distributed through time, the variance of yearly costs $\mathrm{Var}_n^D C_n$ is defined. In Section 5.8, it is going to be minimized to balance the funds spent over time.

It can be proved that minimizing $\mathrm{Var}_n^D C_n$ is the same as minimizing $\sum_n^D C_n^2$.

Proof. In this proof, symbol $\sim$ denotes that two expressions have the same

monotonicity.

$$
\begin{aligned}
\operatorname*{Var}_{n}^{D} C_{n} &= \frac{1}{D} \sum_{n}^{D} \left(C_{n} - \frac{C}{D} \right)^{2} \\
&\sim \sum_{n}^{D} \left(C_{n}^{2} - \frac{2C}{D} C_{n} + \frac{C^{2}}{D^{2}} \right) \\
&= \sum_{n}^{D} C_{n}^{2} - \frac{2C}{D} \sum_{n}^{D} C_{n} + \sum_{n}^{D} \frac{C^{2}}{D^{2}} \\
&= \sum_{n}^{D} C_{n}^{2} - \frac{C^{2}}{D} \\
&\sim \sum_{n}^{D} C_{n}^{2}.
\end{aligned}
$$

Thus, minimizing $\operatorname{Var}_{n}^{D} C_{n}$ is the same as minimizing $\sum_{n}^{D} C_{n}^{2}$. $\qquad\square$

5.7 Constraints to the plan

The only constraint to the plan is due to the limitation of funds. To describe the constraint, the total funds and the cost of the plan needs defining.

Definition 10 (total funds). *The **total funds** is a real number $F \in [0, +\infty)$.*

Definition 11 (cost of project). *The **cost of project** is a real number $c_x \in [0, +\infty)$ defined as*

$$
c_{x} := \sum_{n}^{\infty} c_{x,n}. \tag{7}
$$

The sum c_x must converge because all terms following the d_xth term are 0.

Definition 12 (cost of the plan). *The **cost of the plan** is a real number $C \in [0, +\infty)$ defined as*

$$
C := \sum_{x \in Z} c_{x}. \tag{8}
$$

Due to Assumption 1, the constraint should be

$$
C \leq F. \tag{9}
$$

5.8 The goal of the model

The goal of the model is to

1. Choose proper projects to make the plan have the maximum T,

2. Decide the minimum funds to be raised to satisfy step 1 and the constraint (Equation 9),

3. Decide the start time for each chosen project in step 1 to minimize the variance of yearly costs of the plan.

6 The approach

6.1 Minimize the funds when the time efficiency maximum

7 Pros and cons

Our model have some pros and cons in solving the problem described in 1. The pros are

1. Pro 1.

2. Pro 2.

3. Pro 3.

 The cons are

1. Con 1.

2. Con 2.

3. Con 3.

8   The memo

Memo memo.

9   Appendices

References

[1] Ref 1.