行政院國家科學委員會專題研究計畫　成果報告

# 對動物認知過程做整體推論所必須要的資料型態及分析之探討--以豆象之產卵行為為例

計畫類別：個別型計畫
計畫編號：NSC91-2118-M-004-007-
執行期間：91 年 08 月 01 日至 92 年 07 月 31 日
執行單位：國立政治大學統計學系


計畫主持人：蔡紋琦

計畫參與人員：蔡紋琦、曾奕翔


報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 92 年 10 月 30 日

# 行政院國家科學委員會專題研究計畫成果報告

## 對動物認知過程做整體推論所必須要的資料型態及分析之探討

## --以豆象之產卵行為為例

## What data types and analyses are appropriate for coherent inferences on animal cognitive processing?

## --Illustrated by Female Bean Weevils' Oviposition processes

計畫編號：NSC 91-2118-M-004-007

執行期限：91 年 8 月 1 日至 92 年 7 月 31 日

主持人：蔡紋琦 執行機構及單位名稱:國立政治大學統計學系

共同主持人：無 執行機構及單位名稱：無

計畫參與人員：曾奕翔 執行機構及單位名稱：國立政治大學統計學系

一、中文摘要

很多實驗由於預算上的限制或技術上執行的困難，我們只能在離散的時間點或在某個特定事件發生時的時間點才能對實驗對象做資料的觀察或截取，然後再將手上這份低維結構的資料拿去做分析。但若決定資料取捨的機制和我們所希望下推論的方向有關時，低維結構的資料就可能帶給我們錯誤的訊息，在這種情況下，或許就必須要花費較多的精力去求得高維模組的資料以得到正確的結論。

我們考慮一個由洪和謝所主持的實驗，這個實驗是希望找出母豆象選擇豆子產卵的策略。根據實驗的設計，兩種不同層次的可能模組可被考慮，一個是日模組，另一個是路徑模組。第一個模組不管是從經費或人力資源方面來看，都是比較容易執行取得但其所夾帶的資訊卻少且包含於第二個模組。我們希望藉由電腦模擬去比較是否兩種模組資料都帶給我們一樣好或壞的決策選擇或者是有必要去執行取得第二個較困難但精確的路徑模組。

由電腦模擬的結果，我們發現不管是給定日模組資料或是路徑模組資料，正確的決策都產生較大的最大擬似值，且其和錯誤決策所產生的最大擬似值的比值都差不多一樣大。換言之，低維結構和高維模組兩種資料都給我們一樣好的決策選擇，所以原來根據日模組資料所下的推論是可信賴的，不須要再設法取得更細緻形態的模組資料。

關鍵詞：訊息、模組、策略、最大擬似、認知行為

Abstract

Due to the limitation on budget or technicality, the state of the investigated subject in an experiment usually is recorded only at a discrete time scale or at the time when a certain event occurs. The analysis will then carried out based on the data set of the lower level structure rather than the complete high dimensional temporal platform. If the inference we wish to make is related to the recording switch of a certain event, the lower level structure of data set may conclude incorrect inference and hence it will be worth to make effort on obtaining a higher level of data set.

A simulation is run based on the data collected from an experiment conducted by Horng and Hsieh. In the experiment, they are interested in which strategy that a female bean weevil may use to select their oviposition hosts. Depending on the design of the experiment, two possible levels of data set can be obtained: daily template, or additionally, path template. The daily template is much easier to be observed than

the path template. However it contains less information. The simulation trintends to see whether or not these two different levels of data set bring the same information. If not, a new experiment to obtain a higher level of data set may need to be executed.

The simulation shows that the maximum likelihood value under the true strategy is always larger than the false strategy no matter which level of data set is given and the likelihood ratio magnitude is about the same. This supports us the inference drawn from the lower dimensional structure of data set.

Keywords: Information, Template,
Oviposition Selection Tactic,
Maximum likelihood, Cognition

## 二、Introduction

In many experiments, due to the limitation of the budget or the restriction on the accessibility to the investigated subjects, the data we obtain usually includes only a lower dimensional structure of a high dimensional temporal platform. For instance, the experimenter records the state of the investigated subject only at the end of each day rather than tapes this subject throughout the whole experimental process. Or the data is collected only at the time when a certain button is switched on rather than at the time of each movement of the investigated subject.

It is known that if the lower dimensional structure is a "sufficient statistics" of the high dimensional temporal platform, there will not be loss of information if only the data on the lower dimensional structure is available. However, for a complex system, it is not so easy or obvious to "prove" the sufficiency. Even though, analysis, usually, is still carried out assuming that the lower dimensional structure delivers the same information as the high dimensional temporal platform does. This assumption can be vital. For example, if the recording button is related to the inference we wish to make, we may make even incorrect conclusion with only lower dimensional structure at hand. This project tries to find if a necessary for making effort on obtaining the high dimensional temporal platform to avoid seriously miss-leading by the lower dimensional structure occurs very often. Namely, we wish to see whether or not for most circumstances, it is not that save as we expect to allow the analysis based on only the lower dimensional structure carried on.

We consider an experiment conducted by Hong and Hsieh. The goal of that experiment is finding which strategy a female bean weevil may use to select their oviposition hosts. Each female bean weevil is placed individually into a petri dish containing a certain number of azuki beans with the same size. Depending on the design of the experiment, two possible levels of data set can be obtained. One is the daily record of the egg dispersion at the end of each day during the 8 days of experimental period (daily template), or additionally, the oviposiition state at each time when the female bean weevil lays an egg on a bean (path template). Namely, the first possible experiment produces data of the form $(a(x_1), a(x_2),\ldots, a(x_8))$. The number of eggs that the female bean weevil lays on i-th day is denoted by $x_i$, and the egg dispersion at the end of i-th day is $a(x_i)=(a_0(x_i),a_1(x_i),\ldots)$, where $a_j(x_i)$ is the number of beans which has j eggs at the end of i-th day. The data obtained from second experiment is $(y_1, y_2,\ldots, y_n)$, where $y_i$ is the number of eggs on the bean which the female bean weevil lays her i-th egg on and n is the total number of eggs a female bean weevil lays during the 8 days of experimental period.

The information regarding the oviposition selection tactic brought by the second data set obviously includes the information carried by the first data set. However, there is some restriction in order to execute the second experiment. For instance, the whole oviposition process shall be taped and there shall not be too many beans placed in the petri dish.

## 三、Modeling and Inference

This experiment uses bean with the

same quality (size) and hence assumes that the potential of a bean being an oviposition host is decided by the number of eggs on that bean. The bean with fewer eggs has higher potential. This belief comes from Darwin's "survival of the fittest". Three strategies that the female bean weevils may use to choose their oviposition hosts are considered. The first strategy is the random rule under which the female bean weevil has no preference for any particular bean, no matter how many eggs are on the beans. Second strategy is the absolute rule. This rule assumes the female bean weevil can distinguish the beans having different number of eggs. There is a probability to accept a bean with each level of potential. The probability accepting a bean with i eggs is denoted by p_i, which is assumed to be unknown and decreasing. In other words, the female bean weevil prefers beans having fewer number of eggs. The third strategy is relative rule. This rule assumes that the female has the ability to assess the quality of the whole oviposition environment. In other words, she is able to compare the bean she previously visited with the bean she is standing on now. A previous visit at a worse bean increases the possibility of accepting the current bean as her oviposition host since the environment can be bad if she has visited a bunch of worse beans.

We also assume the time lag between each oviposition is exponential distributed with a location sift. This shift can be explained as the minimum required time to complete an oviposition action. It also helps to decrease the variation while it still keeps the mean oviposition time at a fixed value.

A simulation is run with respect to the absolute rule and the related rule as the true rule. These two rules involve parameters p_0=1>=p_1>=p_2>=… that depend on the bean quality. The true values of the parameters are taken to be p_i=(i+1)^0.504-i^0.504 (Horn 1977). A whole oviposition path is generated based on these two rules respectively and then a sequence of daily egg dispersion data follows. The ML method is used to estimate the unknown parameters and do the strategy selection.

## 四、Simulation Result and Conclusion

The simulation result shows that the true rule always has a higher maximum likelihood value than the false rules (including the random rule) have and the likelihood ratio of the true rule to the false rules is about the same no matter which data set, daily template or path template, is given. The lower dimensional structure and the high dimensional temporal platform both give us a good selection on the oviposition strategy. As to the parameter estimates, both data sets produce very close maximum likelihood estimates. However, the estimates are both far from the true values. This implies the maximum likelihood method might be a bad approach on estimation when this kind of sequential data though it works very well for strategy selection in this model setting.

There is also an issue brought out here. To compute the likelihood value based on the data of daily egg dispersion, we need to exhaust all possible oviposition paths that generate the same daily egg dispersion as the data given. The computation is very huge. One possible way to solve this problem is to select a good representation but small set. From the simulation result, it looks like the likelihood ratio is not affected very much by the paths.

In summary, the simulation result seems to tell us that, under the model setting described in previous section, there is no need to spend more money or time on obtaining the high dimensional temporal platform since the lower dimensional structure data conveys the same information, despite that the inference is correct or incorrect.

## 五、參考文獻


[1] S.B. Horng (1997) Larval Competition and Egg-Laying Decisions by the Bean Weevil, Callosobruchus maculates. Anim. Behav. 53:pp1-12

[2] H.Y. Lin (1997) Statistical Comparison

and Testing on Bean weevil's Egg-Laying Dynamic Rules. Master thesis.

[3] hsieh Fushing, ahwu-Bin Horng, Hui-Ying Lin, and Yen-Chiu Lan (submitted) Testing dynamic rules of animal cognitive processing with longitudinal distribution data