STOPPING RULE DETERMINATION FOR GREEN'S MAXIMUM-LIKELIHOOD ADAPTIVE PROCEDURE WITH PSYCHOGEAR LIBRARY

Marco Vicentini, Diego Dall'Alba, and Debora Botturi
Computer Science Department, University of Verona, Verona, Italy
{vicentini, dallalba, botturi}@metropolis.sci.univr.it

Abstract

This paper presents a case study aimed at identifying the optimal stopping rules for the Greens (1993) Maximum Likelihood adaptive procedure for psychophysics threshold estimation, involving a Monte Carlo simulation. Although this family of adaptive procedure is widely involved in perceptual experiments, there are almost few criteria over a correct and accepted stopping rule. The goal of this work is to identify the optimal stopping rule and its dependence by the minimization function parameters. The simulation runs starting from the capabilities of PsychoGear, a new library of psychophysics methods that supplements visual, audio, and haptics stimuli. The functionalities of PsychoGear let us to easily identify the optimal stopping rule for the Green's procedure. The not trivial Monte Carlo experimental setup is implemented in a clear code which can easily reused in further experiments and simulations.

The study of human perception is expensive and involves a number of pitfalls and difficulties. Over more than a century, methods have been developed and refined that support the sys- tematic exploration within sensory systems of the limits of detection and discrimination among similar and confusable physical stimuli (Leek, 2001).

The characteristics of currently used adaptive procedures are the collection of subject responses to each trial, with a systematic manipulation of the stimulus level along the experimental dimension of interest. Each method results in a series of stimulus levels presented over the course of the experiment, along with the associated subject response (Treutwein, 1995). One family of adaptive procedures has been called the Maximum-Likelihood procedures: their general characteristic is that sets of stimulus-response trials are fit with some type of ogival function; subsequent trial placement and threshold estimation is taken from those fitted functions. This family of adaptive procedures is attractive because it makes full use of all trials in an experiment in order to determine a threshold, rather than estimating threshold only from the levels visited at the end of an adaptive track.

Anyway, there are still some open questions regards the optimal use of these procedure: how is it possible to know how many trials are required by the adaptive procedure to identify the perceptual threshold? Which are the best starting values which allow to minimize the number of trial presentation and assuring a defined accuracy?

The goal of the current work is to identify which is the optimal stopping rule and its dependence by the minimization function parameters for the Green (1993)'s Maximum-Likelihood adaptive procedure. To reach this goal we defined several Monte-Carlo experiments, where the stimuli generation and presentation was managed by PsychoGear, a new library for psychophysics experiment.

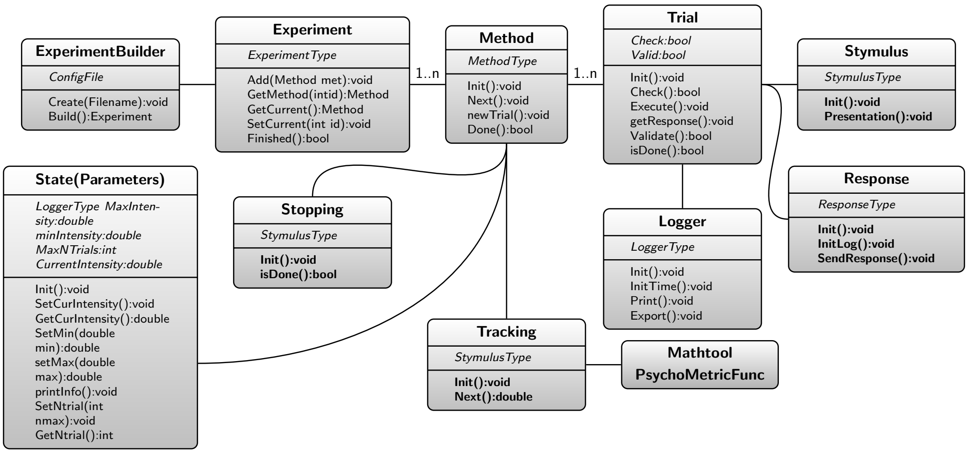


Figure 1: Experimental Psychophysics Library UML Class Diagram.

PsychoGear

Starting from the ideas of the Psychophysics Toolbox (Brainard, 1997) for MATLAB, the PsychoPy–Psychophysics software in Python (Peirce, 2007), and the GroovX framework (Peters, 2008), we are looking at providing new C++ classes and methods to cover several aspects, from stimulus presentation and response collection, using at the same time classical and adaptive procedures, to data analysis, such as psychometric function fitting. Fig. 1 depicts a simplified UML class diagram for the PsychoGear library: each class presents a simple and clear interface, useful to encourage the user with little experience.

The library code is written in C++, is full POSIX compliant, and it works directly under all the common operating systems. The component interface and the code management are very simple, therefore the implementation of new parts is straightforward even for programmers with little experience.

The library architecture is not only due to software engineering, but is meaningful for the organization and the development of every experimental design, in haptic research is even more important, given the complexity of the experiments designs. The basic set of component for a psychophysics paradigm is made of a stopping criterion and an update rule for the characteristics of the stimulus. Hence, if the rules of the experiments are set out, there is no need for complicated control procedure and the execution control could be very easy. Moreover, the possible tracking strategies and stop rules are limited in number, therefore once implemented can be reused.

With this library, once a psychophysics method has been implemented, its reuse in a totally different scenario, with others methods, criteria and devices comes at the cost of the right instantiation of the objects. We plan to provide the implementation of all the most common psychophysics paradigms, the ones discussed in (Leek, 2001). When a experiment design is defined and all the components are implemented, the experimental parameters can be adjusted with no code change or code recompilation. Indeed, an XML configuration file store all the parameters needed for correct execution and so a researcher changes the experimental parameters with a simple text editor.

An experiment built

We employed PsychoGear in a true experimental setup, aimed at identifying the optimal stopping rule for the Green (1993)'s Maximum Likelihood (ML) adaptive psychophysics procedure for threshold estimating, which promises highly efficient trial placement and threshold estimation, by minimizing the number of trials, and then the session duration.

The main feature of any adaptive procedure is that the comparison value which is presented to the subject depends critically on the subject's responses. After stimulus presentation, a set of candidate logistic psychometric functions Φ, defined by a set of parameters Θ, is fitted to all the n data collected up to that point, and the likelihood associated with each function is computed (Treutwein & Strasburger, 1999).

Given the n stimulus presentations at intensities $x_1, \ldots x_n = X$, the likelihood function is defined as

$$\mathcal{L}(\Theta|X) = \prod_{i=1}^{n} L(\Theta|x_i). \tag{1}$$

Each $L(\Theta|x_i)$ is the probability that the subject has given a particular answer – correct or incorrect – when a stimulus with intensity x_i is presented at trial $i.$; it is given by

$$L(\Theta|x_i) = \begin{cases} \Phi(\Theta|F_i) & \textit{if response is correct} \\ 1 - \Phi(\Theta|x_i) & \textit{if response is incorrect} \end{cases} \tag{2}$$

The stimulus level presented on the next trial is the sweet-point of the most likely psychometric function, and it is the stimulus level that minimized the quantity

$$\sigma(x) = \frac{\Phi(\Theta|x)[1 - \Phi(\Theta|x)]}{P'(\Theta|x)^2} \tag{3}$$

where P' is the slope of the function Φ; the minimum value of this last expression, if selected as the stimulus on the next trial, leads to a minimal variability in the estimate of the threshold. The final estimate of threshold is extracted from the most likely psychometric function after some number of trials or when a certain stopping rule is matched.

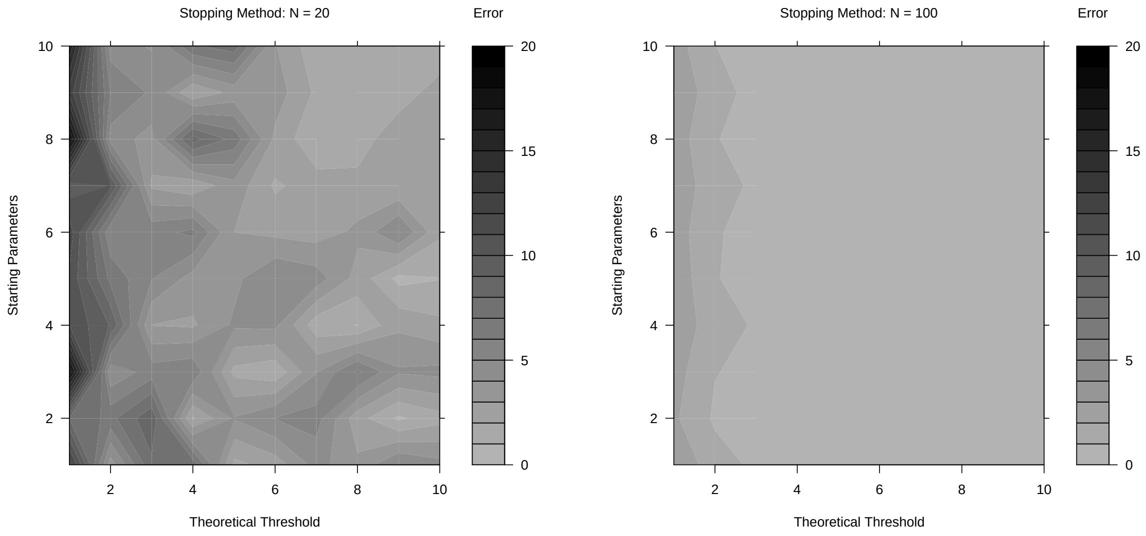
Although this family of adaptive procedure is widely involved in perceptual experiments, there are almost few criteria over the correct stopping rule. Usually, a researcher adopt the strategy to stop when a certain number of trial is proposed to the subject. For example, Green (1993) suggested to stop at 12, 24, or 48 trials, on the basis of the desired accuracy. Elsewhere the strategy is to track the current standard deviation of the estimated parameters until it is lower than a certain desired value. But, how is it possible to know how many trials are required by the adaptive procedure to reach this value?

Moreover, all these criteria depend by the involved minimization function and critically by its starting value. When the starting value is far from the "true" value, it is not guaranteed the algorithm convergency. Which are the best starting values which allow to minimize the number of trial presentation and assuring a certain accuracy?

The goal of the current experiment is to identify which is the optimal stopping rule and its dependence by the minimization function parameters.

Procedure

We conduct Monte Carlo simulations. At first, we define a virtual responder which, according to a Bernoulli random number generator, provides a random response to the



(a) Stopping criterion $n = 20$ (b) Stopping criterion $n = 100$

Figure 2: Percentage error in estimation when the criterion for the stopping rule is the number of trials.

stimuli presented by the adaptive procedure. The responder calculates the probability of the stimulus according to a logistic function characterized by a specific threshold, factorially varied, in the range from 1 to 10. Then, it return the value 1 with success probability p and value 0 with failure probability $q = 1 - p$.

We combine the factor threshold with 10 starting values for the minimization function, also ranging from 1 to 10. Each of the 10×10 combinations is repeated 250 times. We define the stopping value $n = 100$, which represents a criterion double than the maximum proposed by Green (Green, 1993). This experimental setup test 2,500,000 simulated trials.

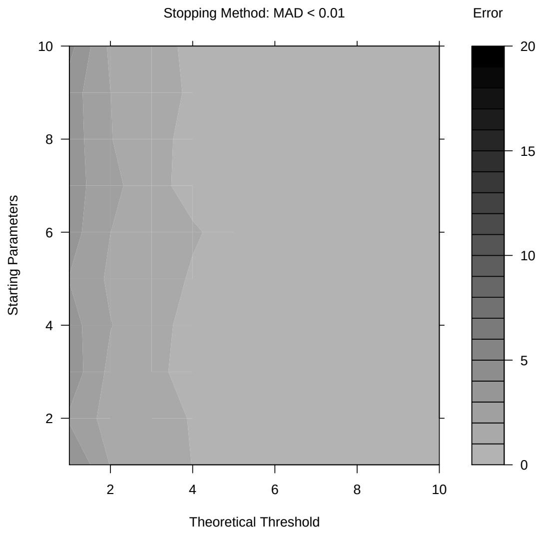
Implementation

We implement the *Tracking* class with the Green's procedure, by defining a ML estimator. This method calls the minimization function from the *MathTool* class and uses the logistic distribution available from the *PsycoMetricFunc* class. The *Response* class was initialized with the just-mentioned virtual responder. The *Stopping* class is set with the value $n = 100$. Then we call the *Experiment* class by defining a full-factorial design with the following parameters: 10 starting values, 10 Bernoulli responders, 250 repetitions.

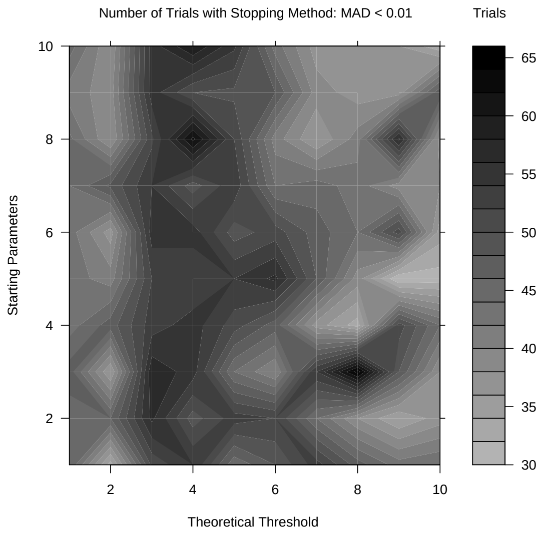
The class *MultiFactExp* extends *Expertiment*: it contains a single *Method* and the command `GetCurrent` sets the correct parameters for the current iteration, while `Finish` verifies if the simulation is completed (250,000 repetitions, each with 100 trials). *MultiFactExp* adds useful parameters to control multi-factorial design, such as current starting value and responder parameter, parameters range, current and maximum number of repetition.

Results

For each trial, we compare the thresholds estimated by our minimization function with the ones asymptotically estimated by the *optim* function implemented within the R statistical



(a) Percentage Error



(b) Trials number

Figure 3: Percentage error in estimation and number of trials according to the criterion for the stopping rule MAD < 0.01

environment (R Development Core Team, 2008). We consider the percentage error in estimation (PE) as goodness of fit index.

As shown in Fig. 2(b), the stopping criterion $n = 100$ assures a strong convergence for all the combination, with a median error of 0.43% (IR between 0.15 and 1.09%). But, if we assume as stopping criterion a shorter number of trials, that is, if we analyze the data by fitting the first $n = 20$ trials, we obtain poor results. Fig. 2(a) shows that, except for a few combinations, the PE is greater than 5% (IR 1.61 to 14.92%), and assumes unsuitable values near 20% for low threshold estimation. Thus, we need to identify the optimal compromise between accuracy, which require a large number of stimulus presentation, and experimental time, related with fatigue and learning effects.

We tested whether the Median Absolute Deviation (MAD) is an efficient rule for stopping the psychophysics procedure. Once, we re-analyze data by considering as stopping rules MAD < 0.1 and MAD < 0.01. While the first rule is totally unsatisfactory (PE = 3.30%, IR from 1.24 to 9.31), the second one results in the best compromise between accuracy and trial number. As shown in Fig. 3(a), the median value of PE is near to the one obtained with $n = 100$: PE = 0.84% (IR ranging from 0.41 to 1.81%). These results are stable across the starting parameters of the minimization function and the responder threshold value. The number of the required trials is generally acceptable, ranging from 35 to 64 (see Fig. 3(b)).

The functionalities of PsychoGear let us to easily identify the optimal stopping rule for the Green's adaptive procedure for psychophysics threshold detection. The not trivial Monte Carlo experimental setup was implemented in a clear code which can easily reused in further perceptual experiments.

Discussion

In this paper we describe how to build a psychophysics experiment with PsychoGear, the library that, taking advantage from the experience of the other implementations and the needs of the haptics perception experiments, proposes an easy to use structure. The novelty of this library is the native management of the psychophysics procedures and of the haptics devices.

The potentialities of PsychoGear are here shown discussing a not trivial psychometrics experimental setup, implemented in a clear code which can easily reused in further perceptual experiments. With our simulations, we identify which we thinks is the optimal stopping rule for the Green (1993)'s adaptive procedure. As also indicated by Wichmann and Hill (2001), we suggest not to count on the trial number, but on the standard deviation of the estimated thresholds. Thus, we propose to involve the Median Absolute Deviation (MAD) as the most efficient rule for stopping the psychophysics procedure, which results as the best compromise between accuracy and trial number.

In the near future we plan to test the capabilities in threshold estimation of all the fundamental adaptive methods. Once the base code allows a fast building process, we will use the current involved procedure as an "how to" for different and more complex Monte Carlo simulations.

We are also planning to to widen this findings also with the support of human subjects in different perceptual context, all broaden by the same psychophysics library.

Acknowledgments

This work was partially founded by the AccuRobAs project under EU's 6th Framework Programme (contract IST-045201).

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433-436.

Green, D. M. (1993). A maximum-likelihood method for estimating thresholds in a yes-no task. *Journal of the Acoustical Society of America*, *93*(4), 2096-2105.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophisics*, *63*(8), 1279-1292.

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8-13.

Peters, R. J. (2008). *The groovx framework v. 1.0a1.* Retrieved July 21st, 2008, from `http://ilab.usc.edu/rjpeters/groovx/`

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org`

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*(17), 2503-2522.

Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophisics*, *61*(1), 87-106.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. fitting, sampling and goodness of fit. *Perception & Psychophisics*, *63*(8), 1293-1313.