

Bayesian Inference and Adaptive Estimation in General Recognition Theory

Joni Pääkkö

Maisterintutkielma
Jyväskylän Yliopisto
MUTKU
Kevät 2020

Abstract

BACKGROUND: Bayesian adaptive estimation has been applied to many problems with good results. However not in this context, no! General Recognition Theory is a multidimensional generalization of Signal Detection Theory.

Aims (which are all intertwined):

- To develop General Recognition Theory by integrating it more with the theory of psychometric function.
 - To argue for abandoning the ubiquitous categorization task, and to instead use a discrimination task (at least in audition)
 - To study YesNo and 2I-4AFC tasks.
- To evaluate the widely used Bayesian adaptive procedure in the context of General Recognition Theory.
 - Does it work? Are there some additional difficulties, or something to take into account?
 - How effective is it? How well are generating parameters recovered? How many trials are needed?
- To facilitate the use of Bayesian inference in GRT.
 - Easy-to-use models are made publicly available
 - Theory is (somewhat) developed

METHOD: Simulation studies and psychophysical studies.

RESULTS: TBA

Contents

1	Simulations	1
1.1	Methods	1
2	Results	1

1 Simulations

Simulations are often used in psychophysics to study the behavior of the algorithms or methods theoretically when analytical solutions are too difficult to obtain. This is the case here too.

There are two main questions:

1. How much more efficient the adaptive algorithm is in relation to sampling stimuli from a fixed grid?
2. How quickly the estimates converge to generating values, and how wide the marginals of the posterior distribution are after N trials?

These questions are related, since *relative efficiency* of the algorithms (question 1) is defined here by the two other quantities: which one of the algorithms manages to more quickly gain information, which is quantified by seeing how quickly the algorithms are able to minimize the marginal standard deviations of the posterior distribution. On the other hand, I'm interested in how quickly the algorithms are able to—or even if they *are* able to—converge to the values of the generating parameters.

The second question has more general value: even if a method would be asymptotically unbiased and consistent¹ usually reducing bias and variance gets exponentially more costly after some point. I'm interested in estimating when this point is reached.

1.1 Methods

Prior distributions Prior distributions and distributions from which generating parameters for the simulations were drawn from are shown in Figure 1. The same prior and generating distribution is used for both dimensions. Note that the scale for criterion is given in false alarm probabilities for easier interpretation.

Priors for the parameters were chosen based on prior information from Silbert, Townsend, and Lentz (2009) and from pilot testing.

Prior for σ was chosen to be fairly vague to reflect the possibility of widely differing thresholds.

Generating parameters for the simulations were drawn from bimodal distributions. The idea was to draw values that are covered by the prior, but which do not necessarily correspond with the mode of the prior distribution. Another motivation was to have qualitatively different simulated observers: some that have high values for some of the parameters and others that have low values.

Parameters in tabular form:

2 Results

¹That is, bias and standard deviation go to zero as N goes towards ∞

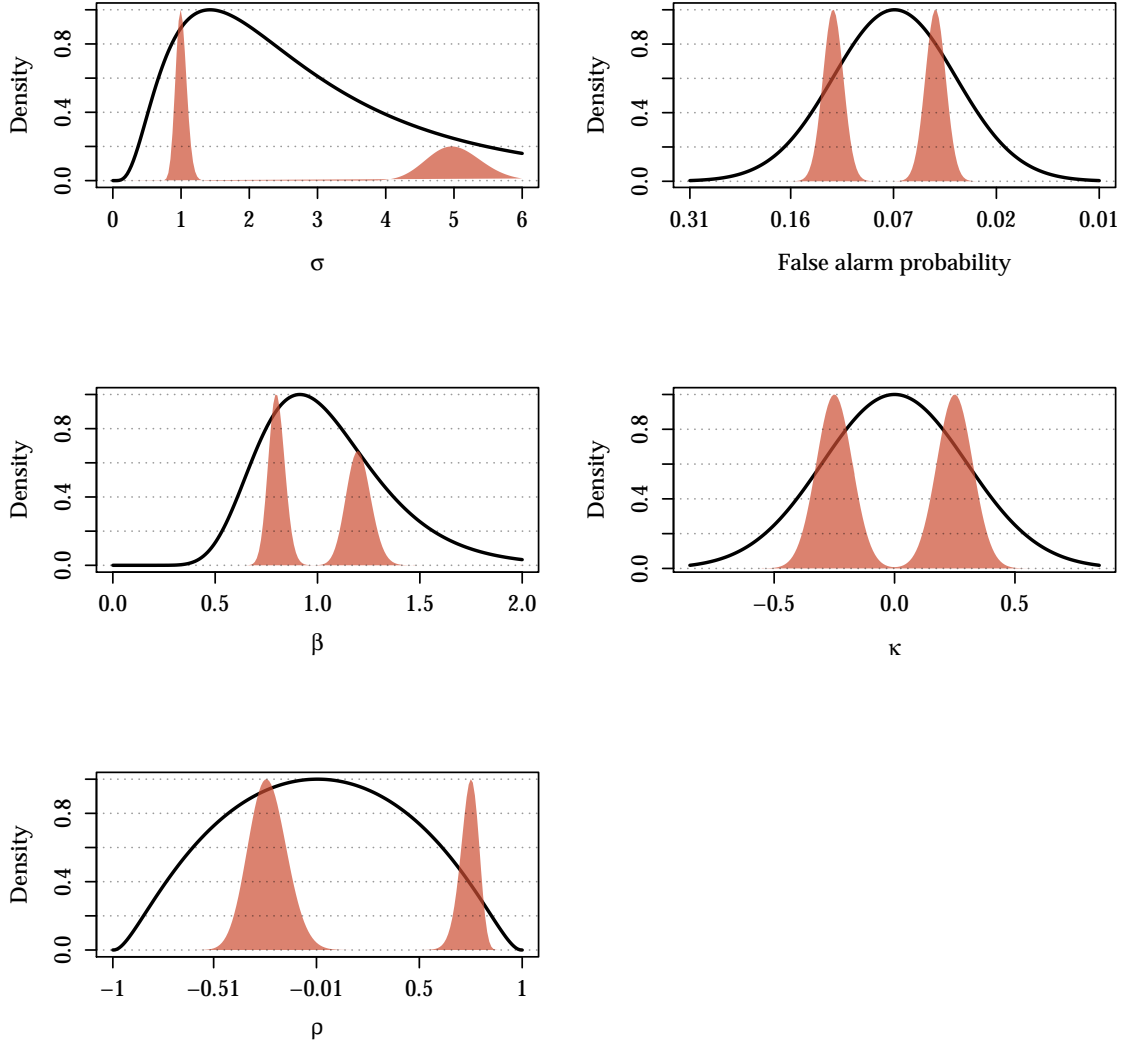


Figure 1: Prior distributions for the parameters of the model (solid black lines) and distributions for generating parameters for the simulations (regions shaded with red). Note that all of the densities are normalized to have maximum value of 1.0.

Table 1: Parameters used for prior distributions and distributions of generating parameters. M^l and M^u , respectively, are for the lower and upper peaks of bimodal distributions.

	Prior		Generating		
	M	SD	M^l	M^u	SD
σ	$\log(2.5)$	0.75	$\log(1)$	$\log(5)$	0.08
C	1.5	0.3	1.2	1.7	0.05
β	$\log(1)$	0.3	$\log(0.8)$	$\log(1.2)$	0.05
κ	0	0.3	-0.25	0.25	0.075
ρ	0	0.7	$\operatorname{atanh}(0.75)$	$\operatorname{atanh}(-0.25)$	0.1

References

Silbert, N., Townsend, J., & Lentz, J. (2009). Independence and separability in the perception of complex nonspeech sounds. *Attention, Perception, & Psychophysics*, 71(8), 1900 - 1915.