

# Likelihood Free Bayesian Adaptive Estimation

Jyväskylän yliopisto  
Musiikin, taiteen ja kulttuurin tutkimuksen laitos  
Pro Gradu -tutkielma  
2025



Joni Pääkkö  
Musiikkitiede  
Marc Thompson

<b>Työn tekijä</b> Joni Pääkkö	
<b>Työn otsikko</b> Likelihood Free Bayesian Adaptive Estimation	
<b>Oppiaine</b> Musiikkitiede	<b>Työn laji</b> Pro Gradu -tutkielma
<b>Aika</b> 5.5.2025	<b>Sivumäärä</b> 85
<b>Tiivistelmä</b> <p>The most popular framework for formalizing adaptive testing – at least in psychophysics – is based on Bayesian statistics and information theory. This framework is usually employed in the context of models that, first of all, contain a likelihood function that can be evaluated in closed form and, secondly, in which responses are discrete (e.g. correct/incorrect).</p> <p>This thesis generalizes the Bayesian adaptive framework to models which lack likelihood function that could be evaluated in closed form due to either its complexity or computational cost.</p> <p>This more general framework can, at least in theory, be applied to a large collection of statistical models: to most, if not all, models used in psychology and related fields. This includes models in which responses are continuous, such as reaction times or ratings on a continuous scale – these kinds of models have presented a challenge for adaptive testing in the past, despite being widely used in cognitive psychology.</p> <p>The thesis is focused on the theory of adaptive testing and does not contain any empirical research or data collection.</p> <p>The overarching structure can be divided in two parts, the first of which explicates and elaborates on the theory of adaptive testing and lays out the mathematical foundations of the titular likelihood free adaptive testing. Second part consists of numerical simulations which are used to evaluate the effectiveness of the proposed adaptive framework and to show how to practically apply it to multiple experimental paradigms that are of relevance for empirical musicological research.</p>	
<b>Asiasanat</b> Adaptive testing, Bayesian statistics, Experimental Design	
<b>Säilytyspaikka</b> Jyväskylän Yliopiston kirjasto	

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Theoretical background</b>	<b>5</b>
2.1	Overview of Adaptive Testing in Psychophysics . . . . .	5
2.2	Bayesian statistics . . . . .	5
2.3	Markov Chain Monte Carlo . . . . .	5
2.4	Likelihood Free Bayesian Estimation . . . . .	5
2.5	Introduction to Bayesian Adaptive Testing . . . . .	5
2.6	Likelihood Free Adaptive Testing . . . . .	5
2.7	Entropy . . . . .	5
<b>3</b>	<b>Black box algorithm</b>	<b>6</b>
<b>4</b>	<b>Simulations</b>	<b>7</b>
4.1	Linear regression with known variance . . . . .	7
<b>5</b>	<b>Conclusion</b>	<b>8</b>

# 1 Introduction

Arguably the most popular framework for formalizing adaptive testing is based on Bayesian statistics and information theory.

## 2 Theoretical background

### 2.1 Overview of Adaptive Testing in Psychophysics

### 2.2 Bayesian statistics

In Bayesian statistics *probability* is defined epistemologically: probability distributions are used to formalize what we *know* of something. Another distinctive feature is that we begin from a formalization of *prior* knowledge that is updated by making observations.

As a simple example we know that in an unmodified deck of cards there's uniform probability for a random card to be any of the four suits. We can represent this with a uniform discrete distribution as in the left panel of Figure 1

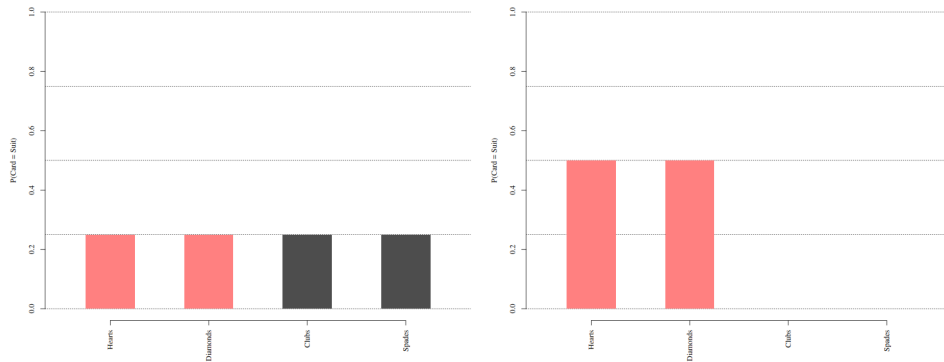


Figure 1:

If our interrogator presents us with the knowledge that the card is red, we can update our knowledge. Now the probability of the card being clubs or spades is zero, and the probability of the card being hearts or diamonds is equal, see the right panel of Figure 1.

### 2.3 Markov Chain Monte Carlo

### 2.4 Likelihood Free Bayesian Estimation

### 2.5 Introduction to Bayesian Adaptive Testing

### 2.6 Likelihood Free Adaptive Testing

### 2.7 Entropy

Entropy of a continuous distribution is

$$h[f(x)] = \int f(x) \log f(x) \quad (1)$$

ERROR REFERENCE

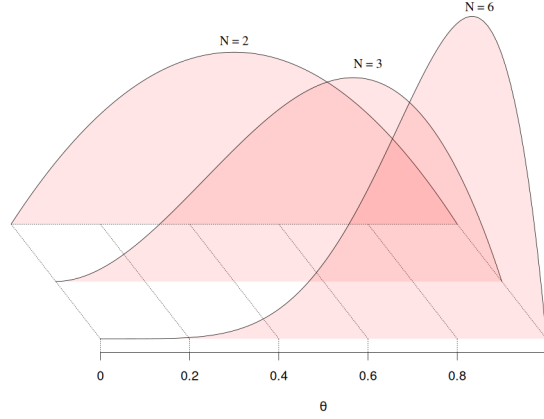


Figure 2: Demonstration of how uncertainty is reduced as more data is observed. Plausible values for the parameter  $\theta$  become more and more concentrated.

In the continuous case we have to integrate over the

$$ig(x) = h \left( \int p(y) \right) - \int h(p(y)) \quad (2)$$

In discrete case the integrals become summations:

$$ig(x) = h \left( \sum p(y)w(x) \right) - \sum h(p(y)w(x)) \quad (3)$$

and, of course, given an IID sample, we can divide the sums with N to get the expectations:

### 3 Black box algorithm

In principle, the proposed framework could be implemented as a black box algorithm that could be used to optimize the data acquisition for any model (as defined in this thesis).

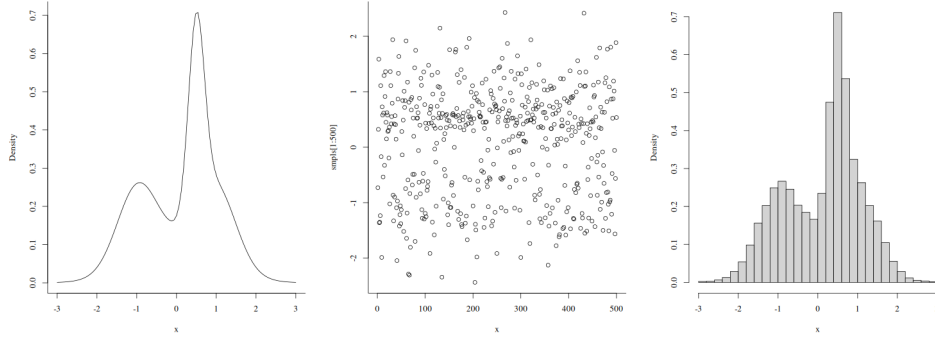


Figure 3:

## 4 Simulations

### 4.1 Linear regression with known variance

When variance is known, the left side reduces to a constant. This does not affect where the maximum of the information gain function is reached. This means that the left side can be dropped.

The equation

$$ig(x) = h\left(\int p(y)\right) - \int h(p(y)) \quad (4)$$

becomes

$$ig(x) = h\left(\int p(y)\right) \quad (5)$$

The distribution of  $p(y)$  is a mixture of normal distributions, and consequently its entropy can not be solved in closed form.

## 5 Conclusion