FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

Pavel Myšička

Analysis of the robustness of the psychometric functions

Department of Software and Computer Science Education

Supervisor of the bachelor thesis:	Filip Děchtěrenko
Study programme:	Computer Science
Study branch:	General Computer Science

Prague 2019

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date .... signature of the author

Title: Analysis of the robustness of the psychometric functions

Author: Pavel Myšička

Department: Department of Software and Computer Science Education

Supervisor: Filip Děchtěrenko, Department of Software and Computer Science Education

Abstract: Psychophysics offers a wide range of experimental techniques to study human perception and often uses mathematical models to do so. Psychometrical function is a formal model of the relationship between intensity of stimulus and perception, that is used by psychophysics to model experimental data. There are various types of psychometric functions used in psychophysical practice. So far it is unknown whether use of different psychometric functions in model experiment data can influence the results of experiment. The goal of this work is to explore the differences between commonly used psychometric functions, prove if there are any differences between their ability to estimate psychometric data and if these differences are big enough so that researchers should pay attention to choice of psychometric function.

Keywords: psychometric function maximum likelihood method robustness

Contents

Introduction

Every science relies on development in it's research. Unfortunately psychological research has been struck by replication crisis several year ago and this situation has shaken the confidence in the whole field. The review of the crisis was sum up in Rouder and Morey [2009]. In response research community is looking for more rigorous and reliable research approaches. Example of some of these approaches can founc in Smith and Little [2018]. Fortunately psychology has entire branch of study that has developed it's experimental designs for more than a century - psychophysics.

Psychophysical experiments are based on simple concept. They consists of stimulus that is presented to the participant and simple, quick and repeatable task, that tests, how is participant able to perceive stimulus. Participant performs the task many times with various intensities of stimulus. Thanks to this straightforward approach, psychophysics can offer experimental designs that are robust, self-replicating, and as simple as possible. With such properties can psychophysical experimental designs produce reliable data even with small number of participants. Psychophysics is not probably only possible remedy for replication crisis, but it is definitely a cheaper and more practical alternative compared to present drastical enlargment of number of participants.

As it was mentioned above psychophysics has developed it's experimental designs for more than century, but that does not mean that such designs are invariable and perfect. There are lots of questions considering experimental design that remain unanswered. We have chosen to tackle the question about role of psychometric functions in experimental design. Psychometric function is used to model the relationship between participant's response and quantified quality of stimulus. It is used as a key model in several experimental designs (methods) of psychophysics. There exist many psychometrical functions, but there is no key or guideline which psychometric function is suitable for which task. There are many questions concerning abilities of psychometric functions, like following. Do particular psychometric functions differ in modeling noised, sparse, or differently imperfect data? Which functions can more easily approximate specific parameters? Or is choice of psychometric functions irrelevant? Answers to these questions are important to every researcher, who has devises his own experiment modeled by psychometric function. Unfortunately today researcher has no assurance, whether the choice of psychometric function is important for experiment or irrelevant. Therefore particular function is chosen by convention, compatibility to other experiments, or personal preference and instinct. These reasons are definitely important and some are very reasonable. But unfortunately this blind spot can still weaken any experiment design using psychometric function and so it is important to this problem to be resolved.

Following thesis is a brief summary of basic psychophysical theory, methodology and describing experiment examining following example of Felix Wichman, Jeremy Hill and other researchers.

0.1 Illustrations

All illustrations and graphics are created in R environment using package gglot2 and in case of additional graphic augmentation are made in io [2017].

0.2 Software

A part of the work contains a simulation experiment conducted in R environment. The experiment uses R package PsyFuns and scripts defining the experiment. Package PsyFuns and all accompanying scripts can be found in osf.io repository.

1. Psychophysics and Psychophysical experiments

Psychophysics is originally science studying relation between physical stimulus and it's psychological perception. During many years it has created comprehensive methodology to measure this relationship. Rather confusingly is this methodology today also referred to as psychophysics, though it is used to study various phenomena beside psychophysic's original field of study. In all respects psychophysics remains interesting field of study and moreover it offers valuable methods and insights to psychology and neuroscience in general Read [2015].

1.1 What is psychophysics

The term psychophysics was used by Gustav Theodor Fechner in book ***Elemente der Psychofyzik***. In this book he has introduced basic principles and methods how to measure mental events and helped to establish psychology as standalone science.

Psychophysics is decribed by Fechner [1860] in his book

Psychophysics studies the properties of sensory apparatus (system). The fundamental characteristics of sensory system according to classical psychophysics are how intense stimuli it can detect and differentiate between stimuli of various qualities like intensity, quality, duration, or combination of all above.

The psychophysics is science field that has evolved since 1880 as described in overview article from Read [2015]. The modern and classical psychophysics view sensory system differently and ask different questions about it. Main concern of classical psychophysics is the *scale of sensation*, which is translation between physical stimulus intensity and level of arousal that this stimulus produces. On the other hand modern psychophysics tends to ask how particular sensory system encodes physical intensity of the stimulus. If we compare these approaches we can see, that modern psychophysics has generalized the subject of it's studies. That means that interpretation of data is now more open towards various models and mechanisms, other than comparing scales of sensation and physical stimulus. But that does not mean that modern psychophysics cannot learn and take inspiration from it's classical counterpart Falmagne [2002].

1.2 Psychophysical experiment

Psychophysical experiment was constructed as a central tool of psychophysics to measure characteristics of sensory systems. It has evolved along psychophysics and today it measures various participant's sensory characteristics using even more various range of techniques. As was mentioned in the introduction - psychophysical experiment consists of stimuli that are presented to the participant and task, that tests participant's ability to perceive stimulus. Participant is presented with various intensities of stimulus and tested by the task multiple times.

The more general overview of psychophysical experiment can be found in Prins et al. [2016].

Researchers have developed various techniques that determine how stimuli will be presented, how will participant will be tested or how resulting data will be processed. Substantial proportion of these techniques uses **psychometric function** as key concept for estimation of characteristics. Remaining techniques generally use averaging of experiment results to estimate resulting characteristics.

As was already mentioned richness and variety of experiment designs in psychophysics is vast.

1.3 Psychophysical experiment example

As an example of psychophysical experiment we will use measuring hearing - threshold, that is quite common experiment. It can be conducted for various reasons from purely medical or scientific to commercial. Our hypothetical example of such reason can be a toy factory trying to make children toys that are "adult-friendly". We all know how annoying can a child with a loud toy be, but what if the toy factory found a way how to make products that are still popular with children and less irritating for adults. We know that children have generally better hearing than adults. So what if toy factory made toys loud enough so toys would be still interesting for children but not yet irritating for adults? To research this option, we need to be able to precisely measure how both children's and adults ability to hear develops with increasing sound intensity. Then we can compare resulting hearing ranges with each other and find optimal intensity of toy sound. Fortunately psychophysics offers us many techniques how to measure hearing ability.

So we conduct an experiment in order to extrapolate children and adult ability to hear sound based on it's intensity. We sample several intensities of sound, let participants listen to them. We would ask participants whether they have heard the sound and calculate their average success rate (performance) per one sound intensity. We can see the performances recorded on particular stimuli intensities as the circles in 1.1.

Data resulting from the experiment are the sets of stimulus intensities and their corresponding performances for each individual - red for an adult and green for a child. It is an example of concept known as small-N design often used in psychophysics Smith and Little [2018]. To represent trend in more general fashion we use the psychometric function to model the data. Psychometric function is a formula that associates intensity of stimulus and performance based on several parameters. By modeling data with psychometric function we have acquired a notion how participants ability to perceive stimulus develops with stimulus intensity.

In next step we have to find way how to compare participants between each other. We can simply compare function's parameters or we can extrapolate more abstract characteristics of participant's performance. These characteristics are called **threshold**, **slope**, etc. and they help to quantify differences between performances of different participants. In case of our hypothetical experiment we can say that child has a better hearing, because it's hearing threshold is

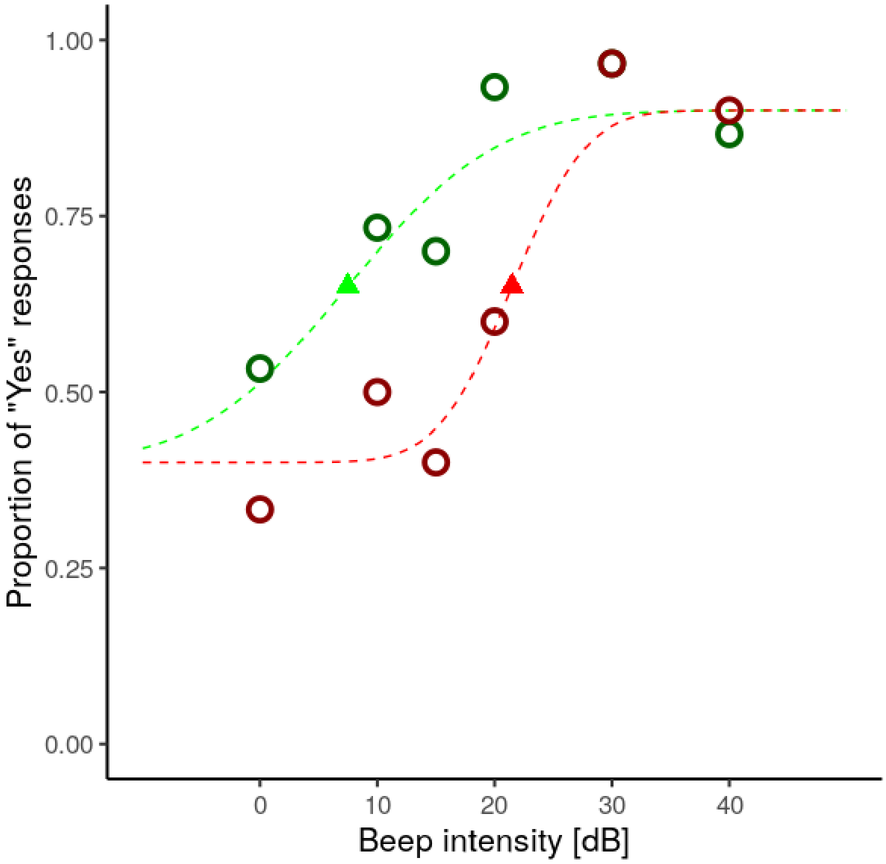

Figure 1.1: In this picture we can see data of two participants of your hypothetical experiment. Participants are divided by color - green are data of child red data of adult. Circles show the performances measured from hypothetical participants. Participants were presented with stimuli of 6 intensities from 0 (no stimulus) to maximal intensity. Each level have been presented 80 times. The lines show the psychometric functions that fit the data and the triangles mark participant's thresholds.

about 15 dB lower, than adult's threshold. Generally have characteristics of psychometric function to be derived from psychometric function parameters, but several functions are designed in such way that particular function characteristics are described by function parameters. We will describe common characteristics of psychometric function in more detail in the next chapter.

1.4 Experiment types and dichotomies

Preceding experiment description and example offer just a small glimpse into the the rich world of psychophysical experiment domain. There are many interesting dichotomies of psychophysical experiment, that provide many nuances to research process. However majority of them does not influence role of psychometric function in the experiment in any significant way. From the experiment properties that effect role of psychometric function in experiment we will name two.

Firstly we will divided experiments to family of ***Method of constant stimuli*** or ***Adaptive methods***. Second division that is important is division to forced-choice Yes-No or Alternative choice design. This Yes-No and Alternative choice design heavily influences sampling scheme used in experiment and therefore is important for psychometric function.

1.5 Method of constant stimuli

Experiment from *Method of constant stimuli* family share approach towards distribution of stimulus and task. This approach is plain and simple - constant stimuli experiments fix the stimuli intensities and number of tasks. Both distribution of stimuli intensities and number of tasks are fixed before the beginning of experiment and are the same for all participants. Only parameter that can vary is order of tasks, that is often randomized. The example from the beginning of the chapter is typical experiment of Constant stimuli method. In case of this experiment researcher decided to use 6 levels of stimuli and also assigned the number of tasks 80 per one level. The set of chosen stimuli levels and the number of observations/samples per level is also called **sampling scheme** and it is an important property of experiment. Results obtained from constant stimuli experiment are predominantly simulated by psychometric function. Because of this close connection is this experimental approach implemented in our simulations.

1.6 Adaptive methods

Adaptive methods are an alternative to Method of constant stimuli. It is a group of experiment methods, that do not fix the intensities of presented stimuli or number of tasks used. Instead they adapt these properties during experiment based on participants responses. Most of these methods are very suitable and effective for measuring single characteristics like hearing threshold Leek [2001]. Only a few can estimate several characteristics at the same time, or approach more complex characteristics. Psychometric functions are used only by a few of adaptive methods. Such methods fit psychometric function with every new response and set stimulus intensity of next task so that they can achieve most information gain. Psi-method is a typical example Prins [2013]. Adaptive methods are not used in our simulations.

1.7 Yes-No / AFC designs

The other parameter of experiment that influences the choice between Yes-No or AFC (Alternative forced-choice) design of task. Yes-No design is older and suitable for measuring detection threshold. That task in Yes-No design is simply to answer whether participant perceived the stimulus - ”Yes” or ”No”. Such approach is simple, but it is hardly applicable for measuring difference thresholds. It is also prone to bias, because participant can incline to answering specifically ”Yes” or ”No” when unsure. The experiment used as example in beginning of the chapter is typical example of Yes-No design.

Alternative to Yes-No design is Alternative forced-choice (AFC). In experiment using this design is participant given several (usually two) alternatives to choose from. According to principles of psychophysics this task is ought to be simple like choosing which of presented pictures contains specific object, which two pictures if is brighter, or which of two beeps is has higher pitch. AFC design experiments can be used for measuring both absolute and difference thresholds. AFC is told to have an advantage to Yes-No design in being bias free. Bias in this context mean a participant's tendency to answer in some specific fashion when unsure. In Yes-No design can participant answer any he wants, when unsure. But actually being unsure means that stimulus was near threshold. If participant decides always to answer no, when unsure, measured data will be skewed upwards from actual threshold. But in AFC design has participant is chance of answering correctly is still $\frac{1}{Numberofalternatives}$ even if participant is prone to choose specific alternative. Though validity of this statement has been argued Klein [2001].

The difference between Yes-No and AFC design is introduced because it can influence the sampling of stimulus intensities that will be presented to the participant. Yes-No design sampling schemes are meant to be symmetrical and widely spread out. In contrast to that sampling schemes used for AFC experiments are rarely symmetrical and are usually skewed towards high performance values.

1.8 Other dichotomies

There are many types of psychophysical experiments, that can be divided into many categories. Experiments can be *Bias-free* or *Bias-dependent*, *Forced-choice* or *Nonforced-choice*, etc. However we must refrain from more detailed description of psychophysical experiments and their properties, because as can be seen they have only limited influence of way psychometric functions and for purpose of this work they would act as distraction. In case of interest more detailed description of psychophysical experiments can be found in the textbook Prins et al. [2016].

2. Properties of psychometric function

Psychometric function is a common way to model and examine relation between stimulus and participant's response and it yields general characteristics of participants performance based on overall results of experiment session. Concept of psychometric function is quite general and it can be applied to many experiment types without any significant alterations. Following chapter should provide the overview over all basic functions properties and some of typically measured characteristics.

2.1 Psychometric function structure

As mentioned above psychometric function is a model of relation between observed participant's performance and power of stimuli presented to observer. Domain of psychometric function depends on nature of stimulus, so generally it is unrestricted, but range of the function is usually the proportion of successfully completed tasks among all tasks therefore it is defined within the interval $(0, 1)$. Psychometric function is assumed to be to be monotonic and increasing. This function properties result in typical "S"-shape of function and such functions are known as **sigmoids**. Generic formula of psychometric function can be described by following equation.

$$\psi(x, \gamma, \lambda, \theta) = \gamma + (1 - \gamma - \lambda) f(x, \phi) \tag{2.1}$$

In following paragraphs we will describe every component psychometric function resulting form previous equation. Scheme of psychometric function structure can be seen in 2.1.

2.1.1 Guess rate γ

Gamma parameter is a lower boundary of observer's performance. Observer cannot detect the signal at this level of performance and is guessing. It is observer's worst possible performance. Example can be 2-AFC design where participant is given two alternatives to choose from. If they guess they have 50% chance of choose correct answer. Similarly in 3-AFC design, where participant is given 3 options, chance of correct guess is 33% and so on. We can see that γ parameter is dependent on **experiment design** and that it is constant during the experiment and among observers. Though observer should not usually score lower than γ, these cases sometimes occur, when observer does not understand task properly, or was insufficiently instructedKlein [2001]. During estimation of parameters guess rate is usually fixed. Estimation of guess rate is conducted only in several experimental designs that can be found in Prins et al. [2016].

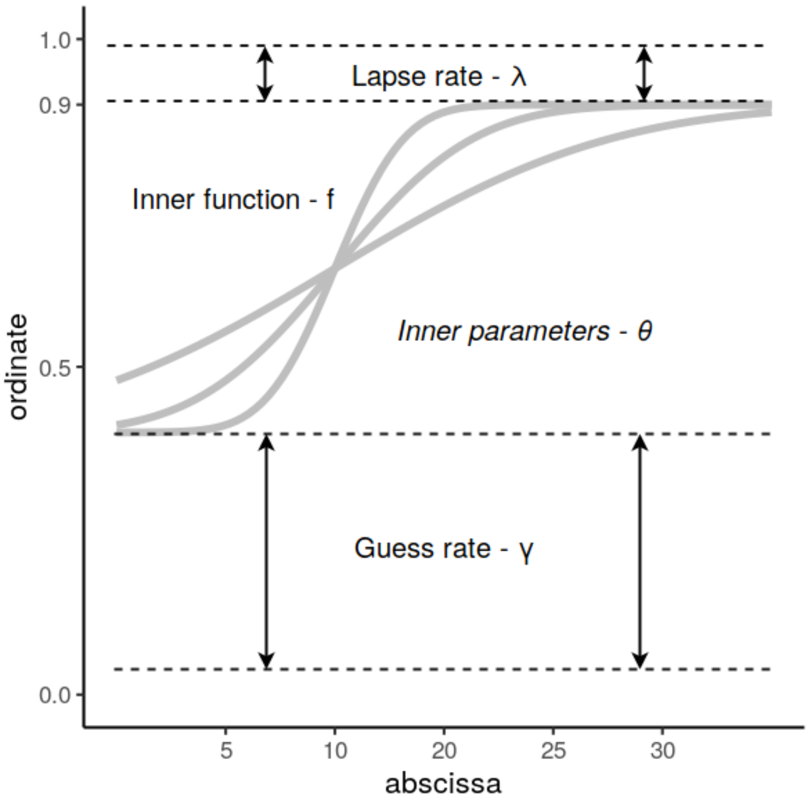

Figure 2.1: In this visualization you can see structure of typical psychometric function. Guess rate γ is a lower boundary of function values. Lapse rate λ is a higher boundary of function values. In between you can see several versions of inner function f based on inner parameters θ.

2.1.2 Lapse rate λ

Lambda parameter determines the upper boundary of performance. This parameter does not depend on experiment design as severely as γ. The most influence is brought by participant and their **lapse rate**. Lapse rate is essentially a number off error observer makes even if he is physically able to detect the signal due to mistake, inattentiveness or lack of motivation. Lapse can vary from observer to observer. Very high lapse rate can damage accuracy of other parameters measurements, so if it exceeds level acceptable by the researcher participant results ought to be removed from experiment data Wichmann and Hill [2001]. Value of high lapse rate can vary by experiment. For 2AFC experiment it are lambda values exceeding 5% considered high lapse rate at which point is participant inattentive every 10th trial.

2.1.3 Inner psychometric function f

Inner function controls the development of performance within the boundaries of guess and lapse rate. Their ability to approximate development of cognitive process is essential to good estimation of threshold and slope parameters.The role of function type in model accuracy is the main concern of this paper. The role of inner psychometric function function have further explained Strasburger [2001], Prins et al. [2016]

Inner function parameters θ

Inner function parameters are unique to any inner function and determine shape of such function. In widely used notation notation they are represented as vector θ and usually no more than two parameter are used. This consensus is kept throughout the range of works Prins et al. [2016], Strasburger [2001], Wichmann and Hill [2001], Klein [2001]. These parameters are not generally interchangeable among different inner functions. Experimenter can calculate measures as threshold and slope on basis of these parameters or he can define function descriptors based on function parameters. There are also many inner functions designed in such way their parameters represent specific kind of function descriptors as threshold. For this reasons θ_1 parameter is often referred to as threshold, because is generally meant to control position of function on abscissa and compared to θ_2 that often influences function's shape and is referred to as slope. In many experiments researchers fix slope parameter for all participants and measure only threshold parameter in order to simplify the experiment example of such experiment can be this work Peng et al. [2013].

2.1.4 Abscissa and ordinate

Abscissa arrangement can also influence shape of psychometric function. There are two widely used arrangements of psychometric functions abscissa - linear an logarithmic. Both stimulus and logarithmic scaling are commonly used and there are no massive differences between their features. Only some measurements of slope have to be adjusting when rescaling. Ordinate is not transformed for classical types of analysis.

In our work we will use term **data** point of describing a point on abscissa.

In experiment using constant stimuli there is clearly defined number and placing of all data points. This arrangement is commmonly called **sampling scheme** and it has been proven to have significant influence on good estimation of psychometric function in several works Lam et al. [1996], Wichmann and Hill [2001].

2.2 Descriptors of psychometric function

Before we start to describe function characteristics we would like to draw readers attention towards difference between function parameters and function characteristics. In overall literature terms threshold and slope refer both to function parameters and to more abstract concepts, that describe properties of sensory systems psychometric functions model. We decided to use term parameters for θ and term function descriptors for more abstract concepts of threshold and slope that will be described further. Other research that have been conducted in this area have had not needed to clarify this division, because it has not needed to compare results of fits between different functions. In situation of experimenting with single function it is completely legitimate to not differentiate between terms descriptors and parameter. However in context of using several psychometric functions at once it is this quite important to mind such division, because parameters are not transferable and comparable among different functions, but descriptors are.

2.2.1 Threshold

Term threshold in this context has a weak or no connection to the term threshold used by classical psychophysics mentioned in first chapter. Threshold is merely a well defined point of psychometric function, so the participants can be compared with each other by single figure. Many psychometric functions are devised in such way so the threshold corresponds with the first of inner parameters θ_1. Literature has so far introduced many types of thresholds and we will list only a most know ones for reference.

Fixed threshold (μ_{fixed})

Fixed threshold is the simplest of kind of threshold. It is merely the defined level of percentage that has to be reached. This type of threshold is not much used as a measure among methods using psychometric functions and it is more popular with adaptive methods not using psychometric functions. The review of such methods can be found in Leek [2001].

Performance threshold (μ_{perf})

Performance threshold is most widely used threshold and most easy to interpret. It is a halfway point of psychometric curve $x = f^{-1}(0.5, \theta)$. This kind of threshold is independent to γ and λ parameter. For many commonly psychometric function (Weibull, Gaussian, Logistic, Cauchy,...) it corresponds with parameter α this fact is also mentioned by Prins et al. [2016].

Improvement threshold (μ_{imp})

Improvement threshold is not as widely used. It is defined as point, where Psychometric function has most rapid growth. Leek [2001]

2.2.2 Slope measurements

The second of inner function parameters is harder to define. The definition of slope is that it should represent steepness of the function. Slope also hints reliability of threshold estimates according to Prins et al. [2016]. Again for the most functions slope is the other parameter of inner function. Slope can be also interpreted as a derivation of psychometric function at threshold Strasburger [2001],Wichmann and Hill [2001]. We have decided to list measurements of threshold that by their definition least closely dependent on function type. They by definition do not use derivations, but quantile distances between function values.

Interquartile range (IQR)

The IQR is defined as the distance of the abscissa between 1st and 3rd quartile of inner psychometric function f. It was introduced as one of measuerements of threshold by Strasburger [2001].

$$IQR = f^{-1}(0.75, \theta) - f^{-1}(0.25, \theta) \tag{2.2}$$

Interquantile range has been proposed as a measure of function steepness along slope, because it can be better compared among different psychometric functions.

Width (w)

The similar term to interquartile range is width. It is based on the same idea but using quantiles 0.95 and 0.05. Schütt et al. [2016]

$$w = f^{-1}(0.95, \theta) - f^{-1}(0.05, \theta) \tag{2.3}$$

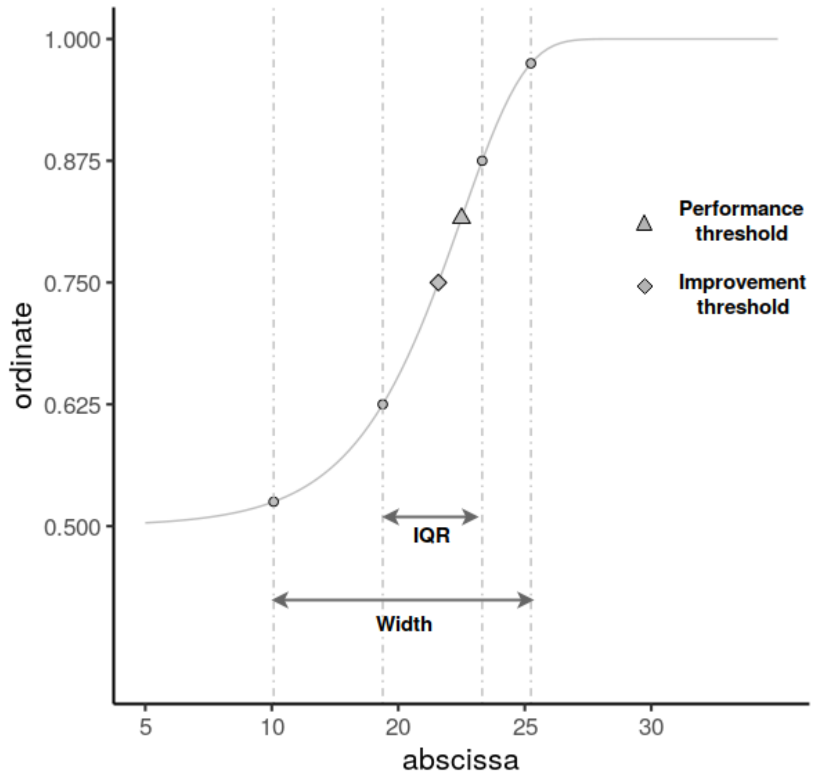

Figure 2.2: In this visualization you can see examples of some psychometric function descriptors. Namely are shown two types of threshold - performance and improvement, and two measures of slope - IQR and width.

2.3 Types of Inner psychometric functions

As was already mentioned psychometric functions are defined on domain in range $(0, \inf)$, or $(-\inf, \inf)$. Ordinate of such function is strictly $[0, 1]$ this has opened the way for all cumulative distribution functions to be used as inner functions of psychometric functions. On the other hand researchers prefer functions with typical "S"-shape to comply with the convention (hence the name sigmoid). Follow examples of typical inner psychometric functions.

2.3.1 Cumulative normal distribution

Cumulative normal is very commonly used probability distribution function with expected value α and variance β. Cumulative normal distribution function can be expressed in an following equation.

$$f_{gauss}(x, \mu, \sigma) = \frac{1}{\beta\sqrt{2\pi}} \int_{-\inf}^{x} e^{-\frac{x-\alpha}{2\beta)^2}} \tag{2.4}$$

In this case is α also value of performance threshold $\mu_p erf = \alpha$ and β is in this case slope parameter.

2.3.2 Logistic function

Logistic is another common inner psychometric function. It can provide good approximation of Cumulative Normal distribution according to Prins et al. [2016] and more so it has closed integral form.

$$f_{logis}(x) = \frac{1}{1 + e^{-\beta(x-\alpha)}} \tag{2.5}$$

In this arrangement α represents performance threshold $\mu_p erf$ and β slope of the function.

2.3.3 Weibull function

Weibull function is one of several functions belongs to the most notoriously known and used functions.. It has been argued by Quick [1974] to be a good approximation for specific types of data.

$$f_{weibull}(x, \lambda) = 1 - e^{-(\frac{x}{\alpha})^\beta} \tag{2.6}$$

It is not symmetric like latter functions and interpretaion of it's parameters is more complicated. Strasburger [2001]. Both function's shape and thresholds are determined by combination of parameters α and β. Therefore conversion is between function parametrs and descriptors of psychometricfunction more complicated than among other functions. Even the simplest traits of function like performance threshold are harder to access $\mu_{perf} = \alpha \sqrt[\beta]{log(2)}$. This makes function's characteristics less clear and accesible.

2.3.4 Left and right Gumbel function

Gumbel function comes from a formula Gumbel distribution used in statistics where it models maximum (minimum) of various distribution samples. This function is also known as Log-Weibull. It was formally introduced among psychometric function by Quick [1974]. Gumbel function is asymmetrical and therefore it can be used in two distinct forms Left Gumbel and Right Gumbel with following equation forms:

$$f_{gumbel-l}(x, \alpha, \beta) = 1 - e^{-e^{\beta(x-\alpha)}} \tag{2.7}$$

$$f_{gumbel-r}(x, \alpha, \beta) = e^{-e^{-\beta(x-\alpha)}} \tag{2.8}$$

Right gumbel is in derived from original "Left gumbel" according to formula $f_{gumbelright}(x, \alpha, \beta) = 1 - f_{gumbelleft}(-x, -\alpha, \beta)$.

Interpretation of parameters is alike to other functions. α is in this case representing threshold and β represents slope. Only in this case performace threshold is acquired by conversion $\mu_{perf} = \alpha + \frac{log(log(2))}{\beta}$ for Left gumbel function and $\mu_{perf} = \alpha - \frac{log(log(2))}{\beta}$ for Right gumbel function.

2.3.5 Quick function

Quick function is closely related to Weibull function and these two functions differ only by base of exponent. However this modification means much more user friendly interpretation of parameters namely performance threshold.

$$f_{quick}(x, \lambda) = 1 - 2^{-(\frac{x}{\alpha})^{\beta}} \tag{2.9}$$

As in Weibull function both shape and threshold of function are determined by combination of inner parameters. However thanks to modification of exponent base is this version of Weibull function performance threshold is always equals to threshold parameter $\mu_{perf} = \alpha$. Prins et al. [2016]

2.3.6 Cauchy function and Hyperbolic Tangent function

As mentioned above any function with sigmoid shape can be used as inner function. The function listed above are among most used and most important. Both following functions are used quite rarely and are listed to demonstrate other possibilities of psychometric function shapes.

Cauchy function is defined as cumulative distribution of the cauchy distribution. Sourceforge in their guidelines suggests, that Cauchy function is more robust towards lapses at high stimulus levels. This idea seems reasonable considering function's shape in high stimulus levels, however we have not been able to find any literature supporting this claim. Strasburger [2001]

$$f_{cauchy}(x, \lambda) = \frac{atan(\frac{x-a}{b})}{\pi} + 0.5 \tag{2.10}$$

Hyperbolic tangent is also quite sparsely used. This function is equivalent to Logistic function with after several conversions, however the relationship between hyperbolic tangent function parameters are similar to Weibull function. The similar shape to Logistic function and it's complicated use of parameters are also the reason of it's rare usage. Strasburger [2001]

$$f_{htan}(x, \lambda) = \frac{1}{2}(1 + \frac{e^{\beta(log(x)-log(a)} - e^{-\beta(log(x)-log(a)}}{e^{\beta(log(x)-log(a)} - e^{-\beta(log(x)+log(a)}} \tag{2.11}$$

2.4 Research of psychometric functions

As was already mentioned research in psychophysics has a long tradition. Researchers have developed improved and compared many techniques of psychophysical experiments. Psychometric function being a key concept for many experimental techniques have also been a subject of many research papers. Reseach concerning psychometric functions has comprised of various topics including anything from theoretical problems like converting between measures of different psychometric measurements Strasburger [2001], or necessary theoretical overviews Klein [2001]. Other theoretical articles tried like Treisman [1999], Green et al. [1966], Garcia-Perez and Alcala-Quintana [2007] try to explain psychometric functions properties according to various theoretical backgrounds and put it in perspective of more complex theories of perception. Other works have focused on practical guidelines concerning methodology of psychometric function usage. One of the most popular and influential paper is the article of Wichman and Hill. It tries to answer the question whether fixing lapse rate is advisable step and under which circumstances it should be used Wichmann and Hill [2001]. There have also been the works that took interest into exploring sampling scheme influence on good estimates of the data Lam et al. [1996]. Unfortunately their results have been argued recently by Prins Prins [2012]. Other works have focused on measurements of goodness of fit Pitt and Myung [2002] or on estimating confidence intervals for function parameters Yi and Merfeld [2016],Schütt et al. [2016] or introduced alternatives to psychometric functions like Model-free estimations Zchaluk and Foster [2009], or further investigation of common practice of fixing psychometric function slope parameter Peng et al. [2013]. However we have not been able to find any work that would concern itself with influence of psychometric function type of function's ability to fit psychophysical data.

2.5 Goals of thesis

Goal of this work and the simulation experiment in chapter 4 is to examine the differences between psychometric functions and their ability to estimate psychophysical data. The experiment is especially dedicated into examining psychometric functions ability to estimate descriptors of psychometric functions as threshold and slope, because these estimates are of most practical use in psychophysical research. The influence of psychometric function will be also studied in relation to used sampling schemes and level of noise added to data.

3. R package PsyFuns

The simulations we are about to perform in this experiment are quite complex and therefore requires tools, that would manage this simulation in such way so it is easy to control and modify. We have decided to execute the simulation in R software environment, because it is widely used by many researchers. However R packages do not provide the support for simulating psychophysical experiment in a way we need. For that reason we have decided to implement our own R package, that would be equally suitable for both simulating and fitting experiment data. The package is called PsyFuns and in following chapter we will try to highlight key ideas behind this package.

3.1 Basic description

PsyFuns R-package was designed in such way so it can both simulate and fit data of psychometric function model. It also implements functions to compute typical descriptors of psychometric functions, like various thresholds and basic measurements assessing goodness of fit. (It contains classes for representation of psychometric function and psychometric function model, which.)

3.2 Representation of psychometric function

The primal purpose for this package is to work with psychometric function therefore the representation of psychometric function in the package are very important. We have decided to follow the model of Python package Psignifit 3.0, that divides representation of psychometric function into sigmoid and core function. This arrangement helps with construction of various psychometric functions and helps to create several variants of the same function. In such arrangement of psychometric function **sigmoid function** provides typical S shape and **core function** scales input parameters. The whole concept is partially resembling structure of generalized linear models as described in McCullagh and Nelder [1989].

Sigmoid

Sigmoid part of inner function is more simple in respect of parameters. It has no parameters adjusting the function shape. Though it has impact on threshold - for instance conversions between some types of thresholds. It provides the typical S-shape. Follows the list sigmoids implemented in PsyFuns.

List of sigmoids

- **gauss** $F_S(x) = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$
- **logistic** $F_S(x) = \frac{1}{1+e^{-x}}$
- **cauchy** $F_S(x) = \frac{tan(x)}{\pi} + 0.5$
- **exponential** $F_S(x) = 1 - e^{-x}$

- **gumbel_l** $F_S(x) = 1 - e^{-e^x}$
- **gumbel_r** $F_S(x) = e^{-e^{-x}}$
- **quick** $F_S(x) = 1 - e^{-x}$
- **htan** $F_S(x) = \frac{tanh(x)}{2} + \frac{1}{2}$

Core

Parameters that can specify inner function's shape apply to core. The structure is somewhat similar to generalized linear models. In following list you can find core functions implemented in PsyFuns.

List of cores

- **ab** $F_c(x, a, b) = \frac{x-a}{b}$
- **linear** $F_c(x, a, b) = ax - b$
- **al** $F_c(x, a, b) = (x - a)b$
- **logy** $F_c(x, a, b) = alog(x) + b$
- **polynom** $F_c(x, a, b) = \frac{x^b}{a}$
- **weibull** $F_c(x, a, b) = 2as\frac{log(x)-log(a)}{log(2)} + log(log(2))$

3.3 Psychometric function model

Psychometric function model is an construct used in PsyFuns package. It is created as a representation of psychometric function, it's parameters, descpriptors of particular psychometric function and table of experiment data in form of data.frame or tibble. Experiment data in form of a table must contain three columns first of them is contains levels of stimuli used in experiment and other two mark number of observations and participant's performance. Alongside compulsory information can table contain any additional information added by user. Psychometric function model can be created by simulation experiment data from a psychometric function, or fitting experiment data.

3.4 Data generation

Simulation of psychophysical experiment data is one of the key parts of PsyFuns package functionality. Therefore package contains class PFm (Psychometric function model) that is designed in such way so that it can create artificial experiemental data based psychometric function is created over and type of noise and noise parameters it is given.

3.4.1 Noise

Package is designed in such way so various types of noise can be used to simulate the data. Current version contains tools to create both additive and multiplicative noise. There are several types of additive noise several from them listed in Petrou and Petrou [2010].

As the additive noise we use symmetrical distributions represented by logistic, normal and uniform distributions. This added by creating noise samples of given mean and standard deviation (in case of uniform distribution it's boundaries) and adding them to previously computed values of psychometric functions. The distribution of noise and value of psychometric function are completely independent of each other resulting in additive noise. On the other hand this can result in combination of psychometric function value and noise fall out of the range [0,1], this is corrected simply by rounding exceeding values to 0, or 1.

3.5 Estimation of psychometric function's parameters

Package was created in such a way so that it is able to both simulate data of psychophysical experiment and estimate psychometric function parameters based on given data. Package uses log-likelihood function and several heuristic in order to estimate goodness of fit of psychometric function parameters on given data and several algorithms to estimate parameters. In following section we will introduce all main components the PsyFuns package uses in order to estimate psychometric data.

3.5.1 Method of maximum likelihood

For estimation of psychometric parameters is generally used method of maximum likelihood. The likelihood function used for psychometrical function has following form.

$$L(\theta, y) = \prod_{i=1}^{N} \left[\binom{n_i}{y_i n_i} \psi(x_i, \theta)^{y_i n_i} (1 - \psi(x_i, \theta))^{n_i(1-y_i)} \right] \tag{3.1}$$

To estimate θ (function parameters), x_i is the stimulus level, n_i is number of observations and y_i is proportion correct. Product $y_i n_i$ gives number of participant's "Yes" answers and vice versa.

Problem of this form of psychometric function is small range of values [0.1] it can reach and therefore simple likelihood limits the precision of algorithms using such values. But logarithm of likelihood can offer much wider range of values $(-\inf, 0]$. Logarithmic function is also monotone and increasing, that means that it can be used instead of simple likelihood function and because of this reason is log-likelihood (logarithmic value of likelihood) much better alternative to plain likelihood function.

$$l(\theta, y) = \sum_{i=1}^{N} \left[log\binom{n_i}{y_i n_i} + y_i n_i log\psi(x_i, \theta) + n_i(1 - y_i) log(1 - \psi(x_i, \theta)) \right] \tag{3.2}$$

Likelihood function that is used for optimizing in the practice slightly modified from the theoretically likelihood. For constant stimuli method the binomial coefficient of likelihood function is always constant and therefore it can be left out. This modification increases precision of computation. This adjustment is also implemented in PsyFuns and other packages working with psychometrical functions like Quickpsy and psyphy. Definite formula of log-likelihood used in package PsyFuns is:

$$l(\theta, y) = \sum_{i=1}^{N} \left[y_i n_i log\psi(x_i, \theta) + n_i(1 - y_i) log(1 - \psi(x_i, \theta)) \right] \tag{3.3}$$

3.5.2 Heuristics

In during the estimation of correct parameters few heuristics are used in order to achieve better parameter convergence. These heuristics come directly from definition of psychometric function and prevent algorithm to pursue estimates that do not comply with this definition. Follows the list of used heuristics and their explanations.

Negative guess or lapse rate Guess and lapse rate cannot be cannot fall under 0, because it would create invalid model. Model constructed with such parameters would assume that participant exceeded 100% or fallen under 0% performance at some point. Therefore are negative guess and lapse rates excluded.

Sum of lapse and guess rate exceeding 1 Such combination of guess and lapse rate is dismissed for same reasons as above.

Decreasing function Inner psychometric function should not be by definition decreasing. If it happens that stimulus decreases the participant's ability to perform given tasks it surely acts as an distractor. Such situation does not occur among usually used experiments designs and therefore it is excluded.

Function midpoint outside the range of sampling scheme Function midpoint is the data point where $f^{(-1)}(0.5)$. It is also the location of performance threshold (2.2.1). Combinations of parameters leading towards this condition are excluded, because there are no commonly used sampling schemes, that would exclude inner function halfway point from their range of values. Therefore such combination of parameters suggests algorithm is converging towards undesired local minimum.

3.5.3 Algorithms

In PsyFuns we have implemented three algorithms for estimating psychometric function parameters.

Default algorithm It the simplest and primal implementation of algorithm for estimating psychometric function parameters. It uses *optim* function for parameter estimation. Parameters of optim function can be adjusted if needed.

Heuristic algorithm Second of implemented algorithms also uses optim function as sub algortihms. However this algorithm estimates psychometric function parameters repeatedly and in each iteration improves chances to find best fit by eliminating local minima. Inner optim function can be also adjusted in case of need.

Evolution algorithm Uses Mullen et al. [2011] implementation of evolution algorithm to find best fit. However this algorithm uses optim function to estimate initial parameters from the data. Yields the stable results but is most time consuming.

We have implemented first two algorithms in such way so they are able to work with multiple version of optim function optimizing algorithms. For simulations used in this work we have chosen Nelder-Mead algorithm, because it has proven most stable in combination with used heuristics. The third algorithm uses Deoptim versions of evolutionary algorithms, but in current implementation it uses Nelder-Mead algorithm for estimating initial parameter if they are not specified.

3.6 Basic package validation

To check basic validity of algorithms for estimating psychometric functions we have conducted several simple tests. We will introduce two of them. In order for our algorithms to be valid they have to yield better fits if presented with data that are more accurate. In case of data from psychophysical experiment this means that algorithms have to estimate underlying psychometric function better once presented with data of better quality (the data that measure performance of participant more precisely). In following simulations we will use two experiments with different accuracy of measuring performance and different levels of noise. We will estimate psychometric functions by all previously mentioned algorithms and measure if they estimate parameters more precisely if given objectively better data set.

3.6.1 Precision of data measurement

In this experiment we want to show that the algorithms are able to estimate parameters from the data more precisely if data are more precisely measured. To do so we have created set of data simulated by closed set of psychometric functions. This set we have sampled with single sampling scheme but with different number of observations per single data point. In this condition better data sets are means data with more observations.

Method

For this simulation we have chosen single inner function. We used Gauss function with parameters $\alpha = 3.5$, $\beta = 1.3$. From one inner function and one set of inner parameters we have created 7 versions of psychometric functions with different lapse rates that has varied between values 0.0, 0.01, 0.02, 0.25, 0.05, 0.075, and 0.1. Guess rate have been kept the same at level 0.5 for all instances of created psychometric functions.

We use one sampling scheme with 320 data points. Sampling scheme is located on interval $(0.5, 6.5)$ and the data points of function are evenly distributed on the interval. For this one sampling scheme we use different number of observations (10,20,40,80,160,320,640,1280,2056,5120,10240,20480,40960). By combining sampling scheme with numbers of observations and psychometric functions we have created set of psychometric function models that we will use to simulate the experimental data. In this instance we will not we will not add any noise to simulated experimental data.

We estimate function's inner parameters and lapse rate based on data for each experiment. Guess rate is kept fixed at 0.5. For estimation of psychometric function we use the same inner function type as was used in data creation.

After estimation of psychometric function parameters we compare goodness of fit of all estimated paramters by comparing all resulting psychometric function with the experiment data and measure their Mean square error.

Results

We compare the resulted MSE measures among all estimated models. The models are divided into groups according to number of observations it was estimated on. The MSE of estimated model should decrease with increasing precision of data values. Number of observations clearly had a positive effect on MSE of estimated psychometric models as we can clearly see in 3.2. And if we conduct ANOVA test for this values we acquire a significant result at the level of significance 0.001 $F(1,271) = 12.99$, $p < 0.001$ - full ANOVA results can be assesed in A.1.

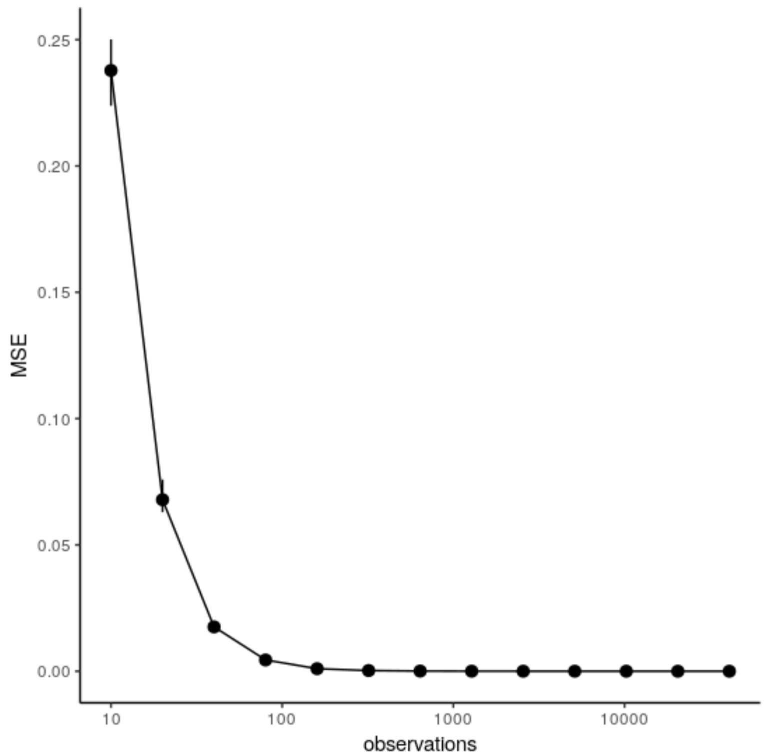

Figure 3.1: In the graph you can see means for values of Mean square errors. The values are divided into groups based on the number of observations original data were conducted over.

3.6.2 Level of noise in the data

In the following experiment is goal to show show that the implemented algorithms are able to fit data more precisely if data contain lower level of noise. To do so we have created set of data simulated by closed set of psychometric function again. This set we have sampled with single sampling scheme and with two different numbers of observations per single data point. Under these condition better data mean data sets with lower noise level.

Method

We used the same function and sampling scheme as used in 3.6.1.

For this one sampling scheme we use two different number of observations 80 and 160. By combining sampling scheme with numbers of observations and psychometric functions we have created set of psychometric function models that we will use to simulate the experimental data. When simulating experiment data we add noise that is normally distributed around original function value. We use 5 different noise distribution with 5 levels of noise deviation 0.01, 0.02, 0.03, 0.04 and 0.05.

The estimation of psychometric function parameters and measures of goodness of fit is again the same as in 3.6.1.

Results

Again we measure MSE of estimated psychometric models. This time resulting MSE is compared based on noise deviation. The higher the noise deviation the worse would lead to higher MSE of estimated model. Noise level had clearly negative effect on mse of estimated psychometric models at significance level 0.001 for five levels of noise $F(1{,}748)=11809$, $p<0.001$ - full anova results can be assesed in A.1. Also we can clearly see this trend in 3.2

3.7 Future development of PsyFuns package

As was already mentioned PsyFuns package was created so that it could offer tools for both fitting and simulating psychometric function models. So far it has implemented several ways hot to represent values of psychophysical experiment, ways to add different types of noise to values of psychometric function and algorithms that estimate function parameters.

This thesis has been first time the package was used in practice. The findings about it's practical use have provided us with many ideas about implementing new and more user friendly functionalities. Moreover we would like to implement wider range of algorithms so we could provide more swift and efficient estimation of psychometric function parameters in the future.

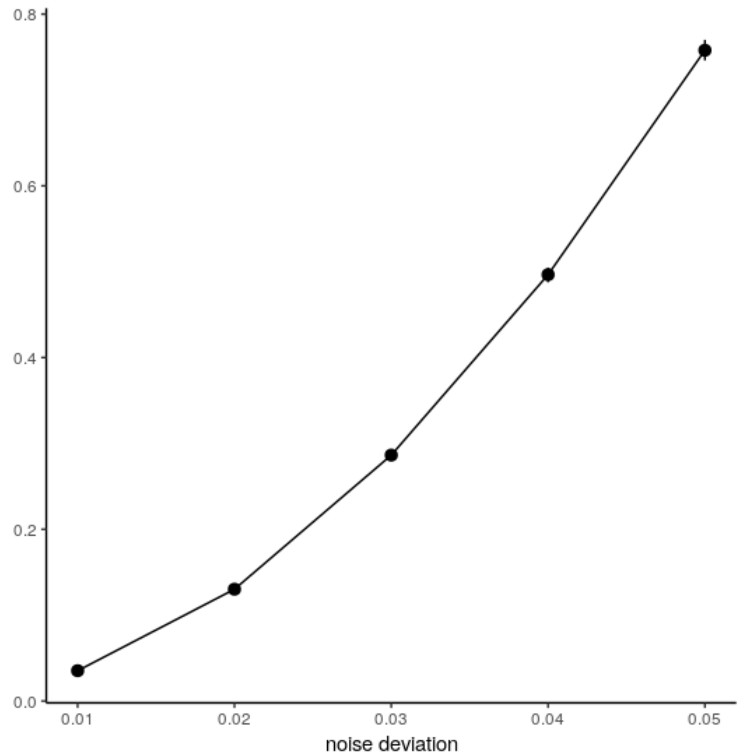

Figure 3.2: In the graph you can see means for values of Mean square errors. The values are divided into groups based on the level of noise that has been added to data.

4. Robustness of usually used psychometric functions

Study of psychometric functions properties is complicated though important part of improving methodology of psychophysics. The easiest and most demonstrative way how to study this property is simulation. Simulation is also easiest and most practical proof of any hypothesis about function's properties. It is not caused by lack of theoretical background, but rather by sole number of influences that can effect the function's ability to deliver desired result.

On the other hand influence of psychometric function on estimates retrieved from the experiment data is rarely mentioned in literature. Even if it is mentioned in any article it does not provide comparison between several functions [Quick, 1974].

The testing of robustness of various psychometric functions is a very complex subject. In order to examine it we will run several simulation experiment to study robustness psychometric function of five different function and analyse the results.

4.1 Method

Simulation experiment concerning psychometrical functions we are about to perform can be divided into two phases - simulation of psychophysical experiments and estimation of psychometric function parameters over simulated data.

First step of experiment is data generation. In this phase are chosen psychometric functions with their parameters. Based on these functions, parameters is simulated a number of artificial experiments. In the following phase of data estimation are these artificial experiments fitted with same functions.

4.1.1 Generation

To generate the data in the first phase of experiment we need to specify types of functions and their parameters. Then we create psychometric function models by combining psychometric functions with sampling schemes and numbers of observations. In the last step of data generation we simulate multiple artificial experiments based on given psychometric function models and noise parameters.

As generating function for final experiment we have chosen 5 typical psychometric functions. We have chosen Weibull function with same parameters as it is used in Wichman-Hill article [Wichmann and Hill, 2001]. Weibull function also served as a specimen and parameters of all other functions were chosen in such a way so all function had alike descriptors. Then we have chosen cumulative normal function and Logistic function as other two commonly used psychometric functions. To accompany the big trio we have chosen Cauchy function. It was chosen from pair of Cauchy and hyperbolic tangent function because hyperbolic tangent's shape is identical with Logistic functions curve and therefore it would not bring any new information. Last of the five function is Left gumbel function. It was chosen from two variants of Gumbel function because it is far more

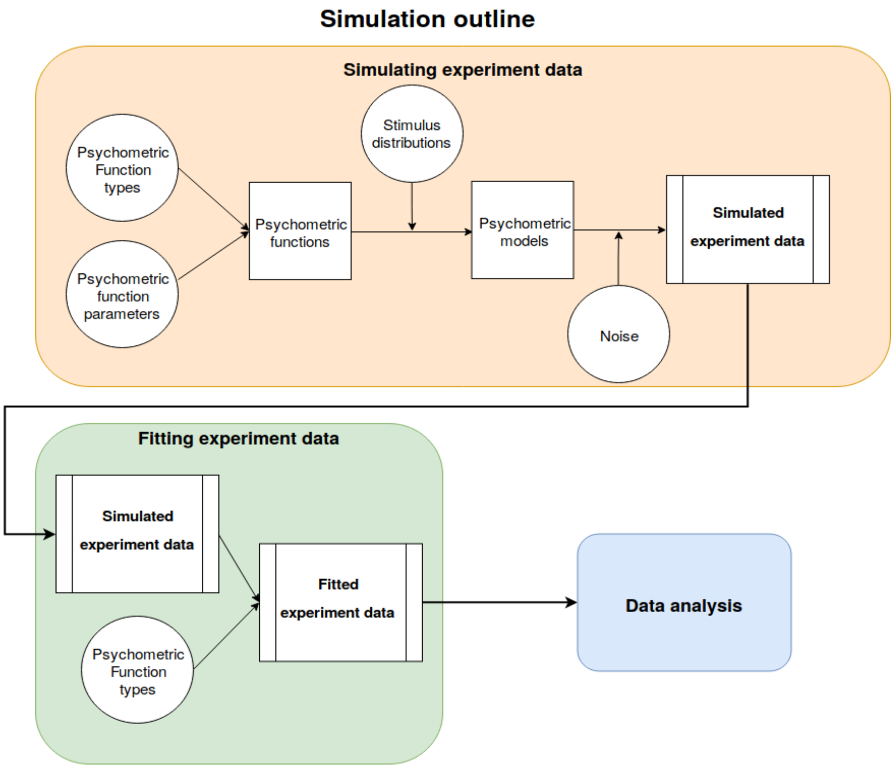

Figure 4.1:

common version of the function. Because there is no need to distinquis it from the Right gumbel function for the rest of the chapter it will only reffered to as Gumbel function.

All inner function parameters are chosen in such way so $\mu_{perf} = 8.85$ and so that IQR approaches value 4.5. Following tables contain the summary of generating functions, their parameters and their characteristics.

To create complete psychometric function from above declared inner functions have to add guess and lapse rates. We set guess rate at level 0.5 for all instances of psychometric functions, but use three different lapse rates 0.03, 0.05, 0.07.

4.1.2 Sampling schemes and observations

Sampling schemes can differ in number and placing of stimulus levels, number of observations per individual level. Wichman has chosen to use 6 samples of

psychometric function	*sigmoid*	*core*	*a*	*b*
Weibull	*exponential*	*polynom*	10	3
Cumulative normal	*gauss*	*ab*	8.85	3.34
Logistic	*logistic*	*ab*	8.85	2.05
Cauchy	*cauchy*	*al*	8.85	0.44
Gumbel	*gumbel_l*	*al*	9.9	0.35

Table 4.1: Table of used psychometric functions, their components and parameters.

stimulus levels and same number of observations, only characteristic that varies is placing of individual levels. Sampling schemes in of psychometric function are chosen based on expected ordinate value of underlying function. This has brought a few complications in our implementation of simulation. In principle it would require each function to have distinct sampling scheme set in the generation process. It would bring another layer of unnecessary complexity to the simulation so we have not decided not to obey this principle and use singular set of sampling schemes for all functions. This set of sampling schemes is prepared for Weibull function used as an example in [Wichmann and Hill, 2001]. The set includes schemes that are located among low (s4) or high (s3,s7) performance values, also distribution contracted around the threshold (s1), or widely dispersed (s5,s2). All sampling schemes are distributed in logarithmic scaling.

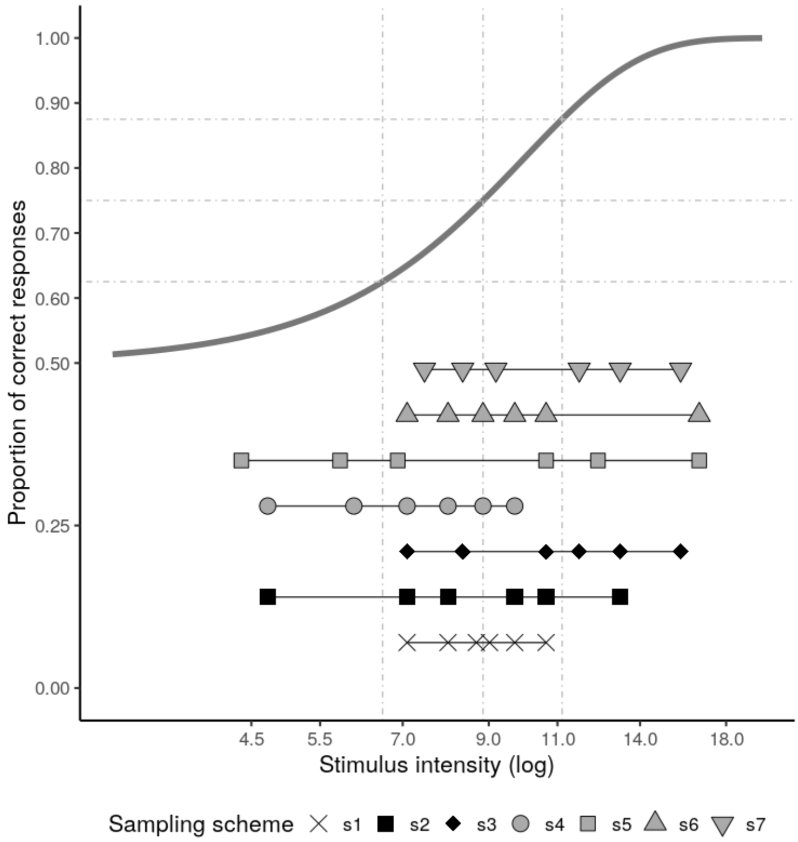

Figure 4.2: Visualization of sampling schemes used in the experiment. Psychometric function above them is Weibull function that shows which values are to be expected on each data point.

Also we have to specify sets observation numbers. We use commonly known numbers of observations - 20, 40, 80, 160 and 240. These are the number of observations that are reasonable to be used in real experiment. The total number of observations per such experiments would be 120, 240, 480, 960 and 1440. Based on given psychometric functions, sampling schemes and numbers of observations we construct set of psychometric functions models that are then used for simulating experiment data.

As function we estimate data from simulated experiments we use again all function types used in generations of data. If we apply these functions on defined sampling schemes, we acquire original psychometric function model. This model yields the same results as would perfect sensory system behaving according to it's associated psychometric function.

4.1.3 Noise

Once we acquire psychometric function models, we can simulate multiple experiments based on these models. However we have to add variability to the simulation in the form of noise. We have decided to use add simple normally distributed noise over the Wichman and Hill version because it can be modulated in easy and understandable way.

4.1.4 Parameter estimation

In second phase we estimate psychometric function parameters form data of simulated experiments with same types of functions as used while generation. We assumed that experiments are conducted in 2-AFC fashion. In such conditions we can fix guess rate at level 0.5. Lapse rate is left as free parameter. For parameter estimation is used PsyFuns heuristic algorithm with optim's Nelder-Mead algorithm [Nelder and Mead, 1965] for inner parameter estimation within individual algorithm steps. After optimal parameters are estimated means of assessing goodness of fit and basic psychometric function descriptors are computed for later analysis.

4.1.5 Descriptors of psychometric function observed in simulation

We have considered several measurements of goodness of fit (AIC, likelihood ratio, MSE, Pearson ^{2}) and descriptors of psychometric functions (μ_{perf},μ_{imp}, IQR, w) to quantify differences between estimations of psychometric functions. Eventually we have chosen estimators over goodness of fit measurements, because they are more interesting from the point of practical use.

Descriptors had to be representing both measures of threshold and slope to describe functions abilities to approximate both these parameters of psychometric function. We have decided to use performance threshold because it is the most used, known and widely understood concept of threshold used in current practice. As a measure of slope we have decided to use interquartile range. We have chosen it over width measure because many functions differ in slope and shape mostly in quite close area around threshold. Therefore IQR represents shape of specific psychometric function more precisely.

In order to analyse functions ability to estimate psychometric function descriptors we have to compute a difference between the original descriptors and descriptors estimated by psychometric functions. We compute Once we compute the difference between original and estimated descriptors and these differences and try to prove if types of psychometric functions have different abilities to do so.

4.1.6 Data analysis

For analysis of our simulation experiment results we use classical statistical methods used in such instance. To find out if there are any significant results within and in between analysed conditions we use two-way ANOVA. If ANOVA shows us sign of difference between psychometric functions ability to estimate descriptors, we quantify the effect sizes with Cohen's d. For classification of Cohen's effect sizes we use rule of thumb suggested in [Cohen, 2013]. The significance of effect sizes is computed by pairwise t-tests using holm method described in [Holm, 1979].

4.2 Results

In following part we will analyse which how do functions differ in ability to fit the data especially what is their ability to to estimate specific descriptors, in this case performance threshold and IQR.

4.2.1 Performance threshold

The performance threshold is one of the most commonly used measures among psychometric functions. Therefore we will analyse it's estimates first. First ANOVA we conduct with factors psychometric function and noise deviance, second with psychometric function and sampling scheme. We will conduct all tests on variable representing performance threshold estimate error so that we can compare functions abilities to estimate performance threshold later.

We conducted two-way analysis of variance on the influence of two independent variables (function, noise deviance) on the performance threshold estimate error. Function type included five levels (Weibull, Cumulative normal, Logistic, Cauchy, Gumbel) and noise deviance consisted of 4 levels (0.1, 0.2, 0.3, 0.4). All effects were statistically significant at the 0.001 significance level for all generating functions, $F(4, 20990) = 89$, $p<0.001$, and for all values of noise deviance $F(1, 20990) = 652$, $p<0.001$. The interaction effect was significant, $F(4, 20990) = 9.70$, $p<0.001$. In A.2 can be found complete data of preceding analysis of variance and in the 4.3 is visualized relationship between psychometric function, noise distribution and goodness of threshold estimation.

Next we wanted to measure interaction between estimating psychometric function and used sampling scheme. Similar ANOVA test we applied to set of factors function and sampling schemes. Function factor included same five types of function and sampling scheme consisted of 7 levels (s1, s2, s3, s4, s5, s6, s7). All effects were statistically significant at same significance level for all psychometric functions, $F(4, 20990) = 86.7$, $p<0.001$, and for all sampling schemes $F(1, 20990) = 19$, p<0.001. The interaction effect between sampling schemes and psychometric functions was significant, $F(4, 20990) = 28$, $p <0.001$. Complete analysis of variance results can be found in A.2 and interaction between factors is visualized in 4.4.

Now we can quantify differences between psychometric functions abilities to estimate the data. To do that we have computed effect sizes pairwise between all functions and the results can be found in 4.2 along with their p-value. Differences between individual function are small according to effect sizes. The only interesting differences are among Gumbel function and the rest. Gumbel function looks that it has no estimated any data well even the data that it has simulated itself. Other differences between psychometric functions are of small effect size or insignificant. We can see that three mostly used psychometric functions Cumulative normal, Logistic and Weibull do not differ among them selves in any interestig way. Other two functions Cauchy and Gumbel yield worse estimates. Cauchy estimates worse than all other functions, but the difference of effect size is fairly small. On the other hand Gumbel function estimates are significantly worse with medium effect size for all other functions.

	Cauchy	Cumulative normal	Gumbel	Logistic
Cumulative normal	0.04***			
Gumbel	−0.43***	−0.46***		
Logistic	0.06***	0.02	0.48***	
Weibull	0.03***	−0.01	0.45***	−0.02

Table 4.2: The table contains pairwise computed differences. The positive value in a row means that function has performed better than one in the column and vice versa. (*** - p <0.001, ** - p <0.01, * - p <0.05, - p <0.05)

4.2.2 IQR

IQR is the second descriptor of psychometric function we want to analyse. However the results are spoiled by Cauchy estimates, that have degenerated and deliver extremely bad estimates for data with noise deviation 0.04. We had to eliminate the data with noise deviance over 0.04 in order to be able to continue with analysis of interquartile range. We again conduct two-way ANOVA's on pairs of factors (psychometric function, noise deviation) and (psychometric function, sampling schemes) and again we use variable interquartile range error in order to be to quantify these differences further.

Two-way ANOVA on the influence of two independent variables (function, noise deviance) on dependent variable interquartile estimate error. Function type included five levels (Weibull, Cumulative normal, Logistic, Cauchy, Gumbel) and noise deviance consisted of 4 levels (0.1, 0.2, 0.3, 0.4). Only the effect of psychometric function was statistically significant at the level 0.001. The main effect for psychometric function yielded an F ratio $F(4, 15740)=715$, $p<0.001$. The main effect for noise deviation resulted in $F(1, 15740)=715$, $p=0.237$. The interaction effect was not significant, $F(4, 15740)=9.70$, $p=0.0804$.

If we apply analysis of variance to pair of psychometric functions and sampling schemes, we obtain more positive results. Effects of both factors are significant. For psychometric function it is $F(4, 15740)=744$, $p<0.001$ and for sampling scheme $F(1, 15740)=42.6$, $p<0.001$. The interaction between variables is also significant $F(4, 15740)=76.3$, $p<0.001$. The full results of ANOVA for both computations can be found in A.2 and A.2. All described relations that influence error of IQR estimate are visualized in 4.5 and 4.6.

According to ANOVA psychometric functions had influence on precision of estimation of interqurartile range. Therefore we can try to compare them and compute effect sizes of these differences. The effect sizes are displayed in 4.3. We can see that results are quite similar. Again Cumulative normal, Logistic and Weibull functions yield better results than other two. Cauchy function has for some reason estimated extremely bad fits for highest level of noise deviation. But for many noise levels below 0.04 it does not seem to give highly worse estimates than other functions. On the other hand Gumbel function has proven to estimate iqr even worse than performance threshold. It significantly worse performance with high effect sizes.

	Cauchy	Cumulative normal	Gumbel	Logistic
Cumulative normal	0.08***			
Gumbel	−0.74***	−0.78***		
Logistic	0.12***	0.02	0.82***	
Weibull	0.02	−0.06***	0.73***	−0.09***

Table 4.3: The table contains pairwise computed differences. The positive value in a row means that function has performed better than one in the column and vice versa. (*** - p <0.001, ** - p <0.01, * - p <0.05, - p <0.05)

4.3 Discussion

From results of ANOVA's and posthoc tests we have obtained an rough image about influence that psychometric functions have on estimating psychometric descriptors. We have proven that there are differences among various psychometric functions in their ability to estimate data of psychophysical experiment at least for this form of simulated data. We have also proven influence of different sampling schemes on both performance threshold and iqr estimates. Wichmann and Hill [2001], Lam et al. [1996]. The influence of noise has been proven only for estimating performance threshold.

The main question that of the whole simulation was if there is a difference between abilities of psychometric functions to fit various data, but the answer is yet ambiguous. In the results we can see that Cumulative normal, Logistic and Weibull functions do not differ in their abilities and always yielded alike results. Therefore these functions seem equally suitable for estimating psychometric data and probably are interchangeable. On the other hand there are Cauchy and Gumbel that have varied form other function in more or less significant way. An important question to answer is what traits shape of psychometric function lead them to yield similar results. This issue might an interesting subject of further work.

But the very important piece of information are results of Gumbel function. In all figures 4.3, 4.4, 4.5, 4.6 it can be seen that is has fitted worse of all functions even for data that were simulated from Gumbel function itself.

We know that there is a significant relation between type of psychometric function, sampling scheme and functions ability to fit both performance threshold and interquartile range. From look at 4.4 we can see that trio of functions Cumulative normal, Logistic, Weibull behave in same way also towards all sampling schemes. They produce best estimates of threshold for sampling schemes s7, s6 and worst for s4 and s1. Sampling schemes s7, s6 have data points in quite long ranges and skewed towards area of high performance compared to sampling schemes s4 and s1. From this two s4 is sampled mainly in low performance data points and s1 has the smallest range of all sampling schemes. So we can probably suppose that trio of functions performs better on spread out sampling schemes that are skewed towards high performance data points. These effects seem to show for both thresholds and iqr estimation.

Cauchy function seems to behave towards sampling schemes in similar way as previous functions, but Gumbel function does not share this tendencies. It's estimates of threshold do deteriorate for sampling schemes s6, s5 and partially for s3. All three sampling schemes contain data points that are among four highest for all sampling schemes. On the other hand does Gumbel function return fairly stable results for sampling scheme s4, that does not contain any high performance data points. This would suggest that Weibull function's ability to estimate thresholds does deteriorate once the sampling scheme contains certain data points in high performance range.

This work is ought to offer guidelines concerning choice of psychometric functions in psychophysical experiments. There are only a few guideline we can give based on our simulation experiment. First advice we can give to researchers that use Cumulative normal, Weibull, or Logistic function in their experiments and are concerned with performance of their functions. We can quite confidently assure them, that they probably do not need to worry about choosing any of these functions. We have proven that Cumulative normal, Weibull and Logistic functions do not differ in their ability to estimate thresholds and slopes in any meaningful way. It is also a positive information for all previous research, because the trio are to most know and used psychometric functions and there interchangeability means that any research done by any of these functions is comparable to any other research done using other two. The second advice we can give to for researchers that for some reason do not use different function than the trio of Weibull, Cumulative normal and Logistic. We can advice to try compare their function to Weibull, Cumulative normal or Logistic function for reference using some standard available tools or package PsyFuns. Because once the function they use proves to be interchangeable with any of the three, it proves that results will be comparable with other research without conversion.

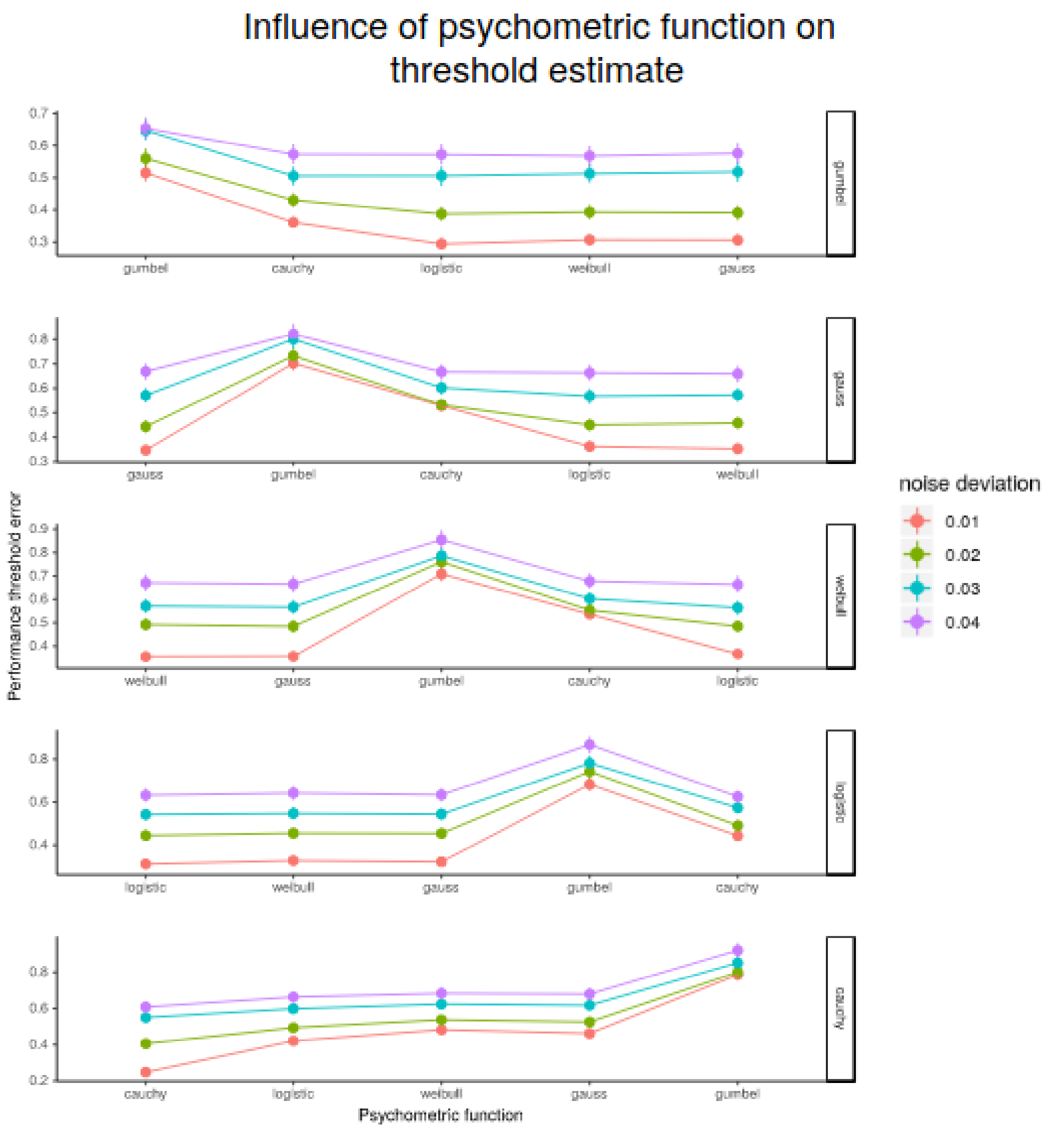

Figure 4.3: In this graph you can see means of differences between original and estimated intequartile range. Results are divided into separated graphs according to functions that have generated the data. The estimating functions can be found on the x-axis and data are further divided by noise deviation level. Every level of noise is assigned a colour.

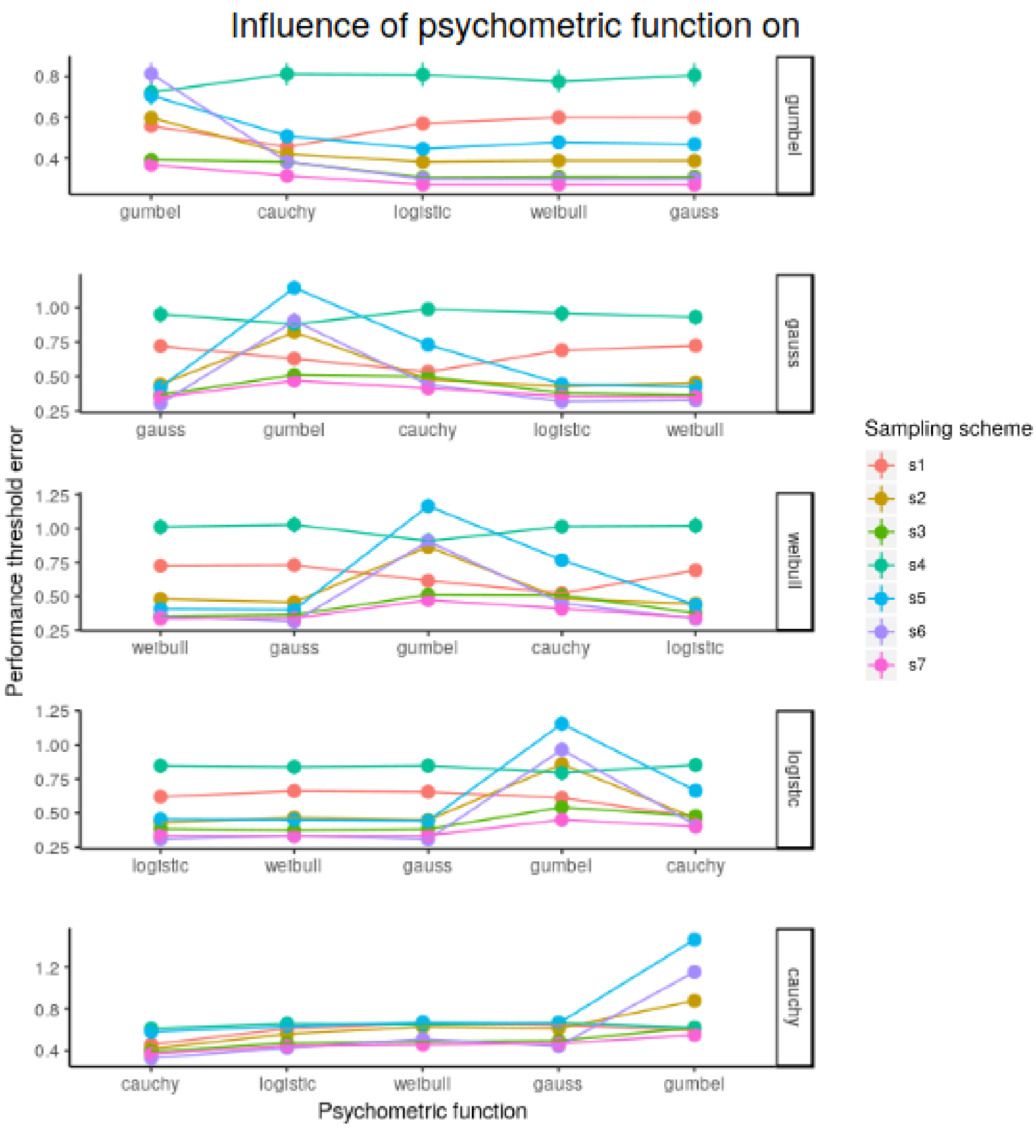

Figure 4.4: In this graph you can see means of differences between original and estimated threshold. Results are divided into separated graphs according to functions that have generated the data. The estimating functions can be found on the x-axis and data are further divided by type of sampling schemes. Every sampling scheme is assigned a colour.

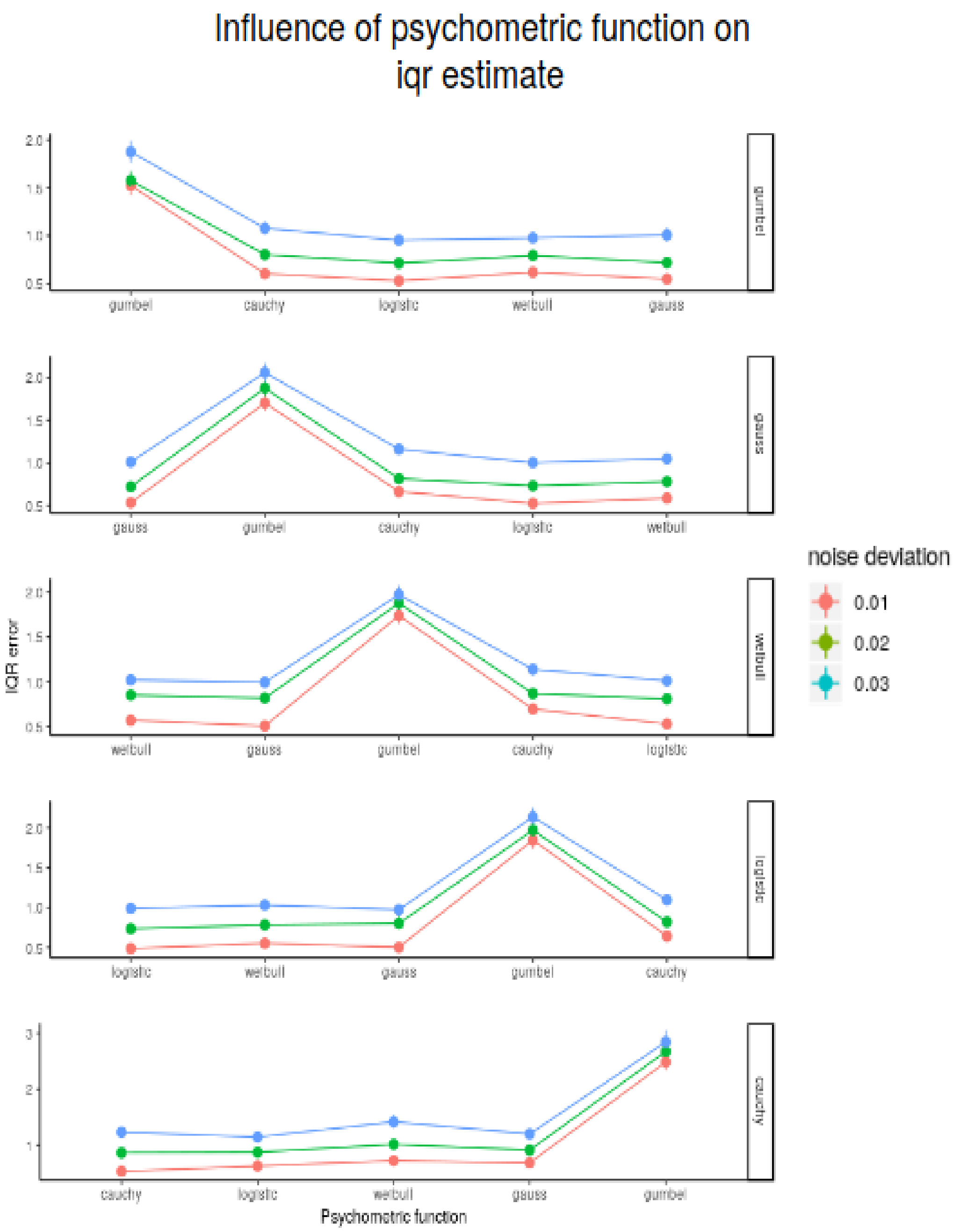

Figure 4.5: In this graph you can see means of differences between original and estimated threshold. Results are divided into separated graphs according to functions that have generated the data. The estimating functions can be found on the x-axis and data are further divided by noise deviation level. Every level of noise is assigned a colour.

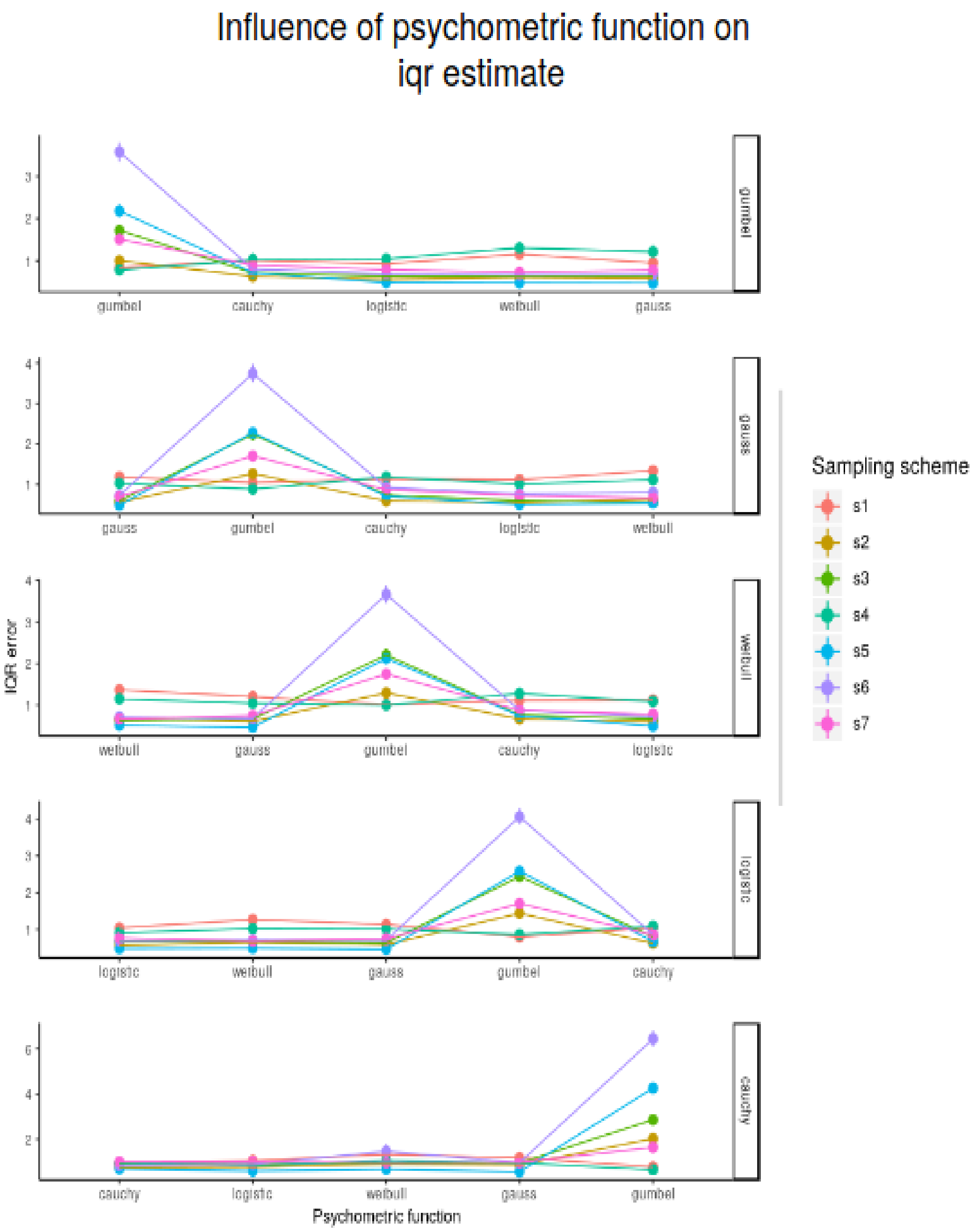

Figure 4.6: In this graph you can see means of differences between original and estimated interquartal range. Results are divided into separated graphs according to functions that have generated the data. The estimating functions can be found on the x-axis and data are further divided by type of sampling schemes. Every sampling scheme is assigned a colour.

Conclusion

In this work we examined psychometric functions, their properties, abilities and role in psychophysical research. The main goal of this work was answering the question whether a type of psychometric function might play a significant role role in functions estimation of psychometric function descriptors as thresholds and slopes.

In first half of the work we have explained the role of psychometric function in the psychophysical research, we have explained a concept of psychometric function and introduced basic function properties and descriptors.

In the second half of the work we have described the package PsyFuns, that was created as a tool for simulating of psychophysical experiments as well as for estimating psychometric function parameters based on these experiments data. Using this tool we have conducted the simulation experiment in order to examine how does the choice of psychometric function influence the estimates of experimental data and how it can estimate psychometric function descriptors as threshold and slope.

We have analysed the data and found significant differences between some functions. These results suggest that choice of psychometric function can have some influence on estimation of descriptors over psychophysical data. However this finding was not found among all studied functions. So far five functions have been examined and only two have differed in any significant way from the group. More so the three functions that were found equivalent are the most used functions of the five that were examined. This leaves the message of our work little ambiguous. On one side it suggests that there might be differences between psychometric functions ability on the other it delivers proof that the most used psychometric functions do not differ in this respect in any meaningful way.

This work has provided only a brief glimpse into issue of psychometric function's role in psychophysical experiment and it has explored this subject only in a fraction of it's complexity. Certainly there is a lot of questions that ought to be answered next including What other factors beside level of noise and sampling scheme influence psychometric function's ability to estimate the descriptors? Are there any rules that could predict functions ability to estimate particular data? How can be psychometric functions compared on real experiment data and can be determined which of them is more suitable? If some of these questions will be answered positively in the future, it could lead the psychophysics to being able to understand this phenomenon and exploit it in order to improve present psychophysical research methods.

Bibliography

Jacob Cohen. *Statistical power analysis for the behavioral sciences.* Routledge, 2013.

Jean-Claude Falmagne. *Elements of psychophysical theory.* Number 6. Oxford University Press on Demand, 2002.

G Th Fechner. *Elemente der psychophysik.* Breitkopf und Härtel, 1860.

Miguel A Garcia-Perez and Rocio Alcala-Quintana. The transducer model for contrast detection and discrimination: Formal relations, implications, and an empirical test. *Spatial Vision*, 20(1):5–43, 2007.

David Marvin Green, John A Swets, et al. *Signal detection theory and psychophysics*, volume 1. Wiley New York, 1966.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

Draw. io. Drawio official website. 2017.

Stanley A Klein. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & psychophysics*, 63(8):1421–1455, 2001.

Chan F Lam, John H Mills, and Judy R Dubno. Placement of observations for the efficient estimation of a psychometric function. *The Journal of the Acoustical Society of America*, 99(6):3689–3693, 1996.

Marjorie R Leek. Adaptive procedures in psychophysical research. *Perception & psychophysics*, 63(8):1279–1292, 2001.

P McCullagh and J Nelder. Generalized linear models second edition chapman & hall, 1989.

Katharine Mullen, David Ardia, David L Gil, Donald Windover, and James Cline. Deoptim: An r package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26, 2011.

John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

Mei Peng, Sara R Jaeger, and Michael J Hautus. Fitting psychometric functions using a fixed-slope parameter: An advanced alternative for estimating odor thresholds with data generated by astm e679. *Chemical senses*, 39(3):229–241, 2013.

Maria MP Petrou and Costas Petrou. *Image processing: the fundamentals.* John Wiley & Sons, 2010.

Mark A Pitt and In Jae Myung. When a good fit can be bad. *Trends in cognitive sciences*, 6(10):421–425, 2002.

Nicolaas Prins. The psychometric function: The lapse rate revisited. *Journal of Vision*, 12(6):25–25, 2012.

Nicolaas Prins. The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of vision*, 13(7): 3–3, 2013.

Nicolaas Prins et al. *Psychophysics: a practical introduction*. Academic Press, 2016.

RF Quick. A vector-magnitude model of contrast detection. *Kybernetik*, 16(2): 65–67, 1974.

JCA Read. The place of human psychophysics in modern neuroscience. *Neuroscience*, 296:116–129, 2015.

Jeffrey N Rouder and Richard D Morey. The nature of psychological thresholds. *Psychological Review*, 116(3):655, 2009.

Heiko H Schütt, Stefan Harmeling, Jakob H Macke, and Felix A Wichmann. Painfree and accurate bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, 122:105–123, 2016.

Philip L Smith and Daniel R Little. Small is beautiful: In defense of the small-n design. *Psychonomic bulletin & review*, 25(6):2083–2101, 2018.

Hans Strasburger. Converting between measures of slope of the psychometric function. *Perception & psychophysics*, 63(8):1348–1355, 2001.

Michel Treisman. There are two types of psychometric function: A theory of cue combination in the processing of complex stimuli with implications for categorical perception. *Journal of Experimental Psychology: General*, 128(4): 517, 1999.

Felix A Wichmann and N Jeremy Hill. The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & psychophysics*, 63(8):1293–1313, 2001.

Yongwoo Yi and Daniel M Merfeld. A quantitative confidence signal detection model: 1. fitting psychometric functions, 2016.

Kamila Zchaluk and David H Foster. Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics*, 71(6):1414–1425, 2009.

List of Figures

List of Tables

A. Attachments

A.1 Full ANOVA tables for tests run on package PsyFuns

Table A.1: Results of single-way ANOVA of MSE and number of observations.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
observations	1	0.05	0.05	12.99	0.0004
Residuals	271	1.09	0.00		

Table A.2: Results of single-way ANOVA of MSE and noise deviation.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
noise_sd	1	49.22	49.22	11808.76	0.0000
Residuals	748	3.12	0.00		

A.2 Full ANOVA tables for simulation experiment results

Results of two-way analysis of variance used in chapter 4 in full extend.

Table A.3: Results of two-way ANOVA for factors of psychometric functions, noise deviance and dependent variable of measuring distance between estimated and performance threshold.

	sigmoid_gen	Sum Sq	Df	F value	Pr(>F)
1	cauchy	350.88	4.00	416.35	0.00
2	cauchy	163.06	1.00	773.94	0.00
3	cauchy	16.56	4.00	19.65	0.00
4	cauchy	4422.32	20990.00		
5	gauss	204.71	4.00	214.81	0.00
6	gauss	173.07	1.00	726.42	0.00
7	gauss	22.47	4.00	23.58	0.00
8	gauss	5000.87	20990.00		
9	gumbel	71.26	4.00	89.04	0.00
10	gumbel	168.02	1.00	839.78	0.00
11	gumbel	7.76	4.00	9.70	0.00
12	gumbel	4199.55	20990.00		
13	logistic	249.70	4.00	287.82	0.00
14	logistic	198.08	1.00	913.28	0.00
15	logistic	11.75	4.00	13.54	0.00
16	logistic	4552.58	20990.00		
17	weibull	209.27	4.00	208.32	0.00
18	weibull	163.64	1.00	651.60	0.00
19	weibull	18.16	4.00	18.08	0.00
20	weibull	5271.29	20990.00		

Table A.4: Results of two-way ANOVA for factors of psychometric functions, noise deviance and dependent variable of measuring distance between estimated and performance threshold.

	sigmoid_gen	Sum Sq	Df	F value	Pr(>F)
1	cauchy	350.88	4.00	405.71	0.00
2	cauchy	4.11	1.00	19.02	0.00
3	cauchy	59.49	4.00	68.79	0.00
4	cauchy	4538.33	20990.00		
5	gauss	204.71	4.00	211.03	0.00
6	gauss	57.74	1.00	238.07	0.00
7	gauss	48.27	4.00	49.76	0.00
8	gauss	5090.40	20990.00		
9	gumbel	71.26	4.00	86.75	0.00
10	gumbel	41.73	1.00	203.21	0.00
11	gumbel	22.97	4.00	27.97	0.00
12	gumbel	4310.63	20990.00		
13	logistic	249.70	4.00	280.13	0.00
14	logistic	42.76	1.00	191.86	0.00
15	logistic	42.02	4.00	47.14	0.00
16	logistic	4677.64	20990.00		
17	weibull	209.27	4.00	205.84	0.00
18	weibull	63.86	1.00	251.23	0.00
19	weibull	54.30	4.00	53.41	0.00
20	weibull	5334.94	20990.00		

Table A.5: Results of two-way analysis of variance for factors of psychometric functions and noise deviance and dependent variable of measuring error in estimating IQR.

	sigmoid_gen	Sum Sq	Df	F value	Pr(>F)
1	cauchy	15528.55	4.00	1104.37	0.00
2	cauchy	4.92	1.00	1.40	0.24
3	cauchy	53.54	4.00	3.81	0.00
4	cauchy	55330.21	15740.00		
5	gauss	6204.63	4.00	849.33	0.00
6	gauss	60.20	1.00	32.96	0.00
7	gauss	48.54	4.00	6.64	0.00
8	gauss	28746.49	15740.00		
9	gumbel	5016.02	4.00	715.00	0.00
10	gumbel	101.66	1.00	57.96	0.00
11	gumbel	26.35	4.00	3.76	0.00
12	gumbel	27605.73	15740.00		
13	logistic	7053.18	4.00	908.22	0.00
14	logistic	10.82	1.00	5.57	0.02
15	logistic	16.16	4.00	2.08	0.08
16	logistic	30559.08	15740.00		
17	weibull	6345.33	4.00	896.08	0.00
18	weibull	109.01	1.00	61.58	0.00
19	weibull	33.11	4.00	4.68	0.00
20	weibull	27864.57	15740.00		

Table A.6: Results of two-way analysis of variance for factors of psychometric functions and noise deviance and dependent variable of measuring error in estimating IQR.

	sigmoid_gen	Sum Sq	Df	F value	Pr(>F)
1	cauchy	15528.55	4.00	1170.58	0.00
2	cauchy	1229.66	1.00	370.78	0.00
3	cauchy	1958.30	4.00	147.62	0.00
4	cauchy	52200.70	15740.00		
5	gauss	6204.63	4.00	877.99	0.00
6	gauss	493.95	1.00	279.59	0.00
7	gauss	553.05	4.00	78.26	0.00
8	gauss	27808.24	15740.00		
9	gumbel	5016.02	4.00	743.63	0.00
10	gumbel	71.79	1.00	42.57	0.00
11	gumbel	1119.25	4.00	165.93	0.00
12	gumbel	26542.71	15740.00		
13	logistic	7053.18	4.00	944.54	0.00
14	logistic	422.55	1.00	226.35	0.00
15	logistic	779.65	4.00	104.41	0.00
16	logistic	29383.87	15740.00		
17	weibull	6345.33	4.00	928.06	0.00
18	weibull	580.59	1.00	339.66	0.00
19	weibull	521.81	4.00	76.32	0.00
20	weibull	26904.30	15740.00		