Spring 5-2020

A Generalized Gaussian Process Likelihood for Psychometric Function Estimation

Jonathan Wenhan Chen
Washington University in St. Louis

Washington University in St. Louis

McKelvey School of Engineering

Department of Computer Science and Engineering

Thesis Examination Committee:
Dennis Barbour, Chair
Netanel Raviv
Brian Garnett

A Generalized Gaussian Process Likelihood for Psychometric Function Estimation

by

Jonathan Wenhan Chen

A thesis presented to the McKelvey School of Engineering
of Washington University in partial fulfillment of the
requirements for the degree of

Master of Science

May 2020
Saint Louis, Missouri

Acknowledgments

I would like to, first, thank my mentor, Associate Professor Dennis Barbour, for introducing me to this subject and supporting my work through his mentorship. Without his guidance, I would not have been able to learn about and work on such an exciting topic and application of machine learning. I would also like to thank Assistant Professor Roman Garnett for helping me debug my Gaussian processes, often late into the night on the weekends, and for coaching me through analyzing them. I cannot thank them both enough for their patience over the last two years.

A special thanks must also go to my lab mates Trevor Larson and Kevin Xie, for their help on this project and throughout my time in Dr. Barbour's lab. Without their support and friendship, the completion of this thesis would have been infinitely more challenging.

I would also like to thank my parents and my friends for supporting me throughout my studies through times of stress and times of happiness. This accomplishment would truly not be possible without them.

Jonathan Wenhan Chen

Washington University in Saint Louis
May 2020

Dedicated to my parents.

Contents

List of Figures

ABSTRACT OF THE THESIS

A Generalized Gaussian Process Likelihood for Psychometric Function Estimation

by

Jonathan Wenhan Chen

Master of Science in Computer Science

Washington University in St. Louis, May 2020

Research Advisor: Associate Professor Dennis Barbour

Psychometric functions model the relationship between a physical phenomenon, an independent variable, and a subject's performance on a cognitive task. The estimation of these psychometric functions is critical for the understanding of perception and cognition as well as for the diagnosis and treatment of many sensory conditions. The ability to estimate psychometric functions of any complexity is necessary to this end. In the following thesis, a generalized likelihood function for psychometric function estimation with Gaussian processes is described and validated. Such a likelihood function is necessary to enable the usage of Gaussian processes for the estimation of non-zero guess and lapse rate psychometric functions. It is also applicable, in general, to any problem where the probability of one or more classes has theoretical non-whole upper or lower asymptotes.

Chapter 1

Introduction

In the study of perception and cognition, one of the primary objectives is to understand a subject's ability to perceive and react to various stimuli. This is typically done by asking a subject to perform a task involving the processing of a stimulus. These tasks can be anything from merely listening for a sound and pressing a button to more complex tasks like solving a puzzle. The critical feature for any of these tasks is that their difficulties can be manipulated. The parameterization of these tasks enables the quantification of a subjects performance and ability to complete varying levels of the task. As a task becomes more difficult, a subject's success rate may decrease until the task cannot be completed. The successful development of subject task performance models may be useful, if not critical, not only for scientific understanding but also for diagnosis (e.g., measuring understanding in testing) and treatment (e.g., the optimal design of hearing aids).

1.1 The Psychometric Function

A psychometric function (PF) relates the parameters of some stimulus to the response that it elicits, typically measured by a subject's performance on a cognitive task [13]. Mathematically, PFs are considered to map some input domain to the success probability of a Bernoulli random variable. Thus, they have historically been assumed to be sigmoidal and are parameterized by a threshold stimulus value α above which responses are probably successful, and a spread value β describing the rate of transition from mostly unsuccessful

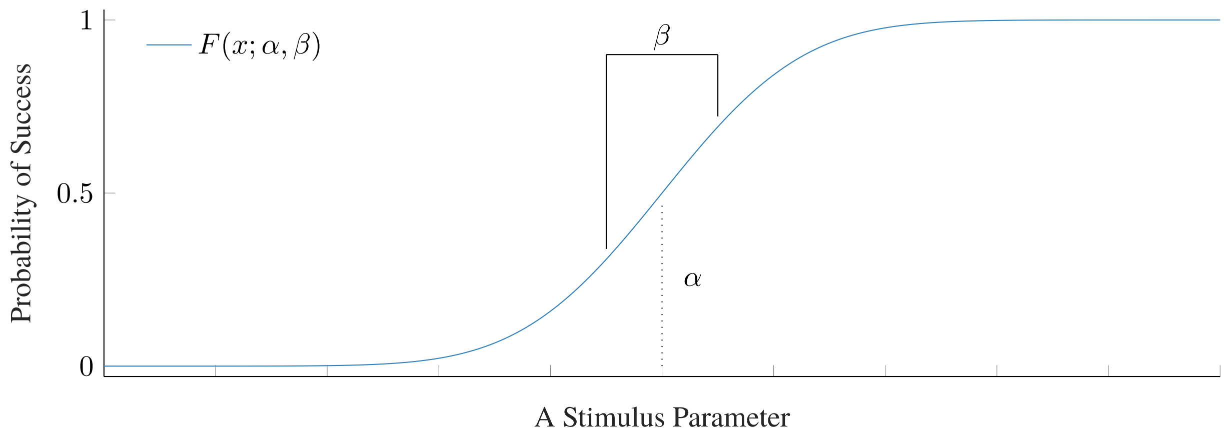

Figure 1.1: A Psychometric Function

responses to mostly successful. In the following sections, the parameters α and β will be collectively referred to as θ. An example of a PF for an arbitrary task is shown in Figure 1.1.

A simple sigmoid is a sufficient model function for PFs only in special cases where guessing, successes in regions where performance is expected to be poor, and lapses, mistakes when performance is expected to be good, are both limited. One such case is the measurement of hearing ability. In the task used, a tone is played for the subject at various intensities or "loudnesses" and a subject is asked to report whether or not they heard it. Since the failures to detect the tone are likely due to an inability to hear the tone and not the inability to process it, the risk of guessing is low. Furthermore, with hearing, if a sound is loud enough to be heard, then there is also little risk of lapse, an unexpected mistake. This is not the case for many other tasks. For any task framed where a subject is required to choose between several options. In these tasks, there is a theoretical guess rate as the subject is forced to randomly choose one of the choices if they do not know the right answer. To take both guesses and lapses into account, a generalized PF given is by

$$\psi(x; \gamma, \lambda, \theta) = \gamma + (1 - \gamma - \lambda)F(x; \theta), \tag{1.1}$$

where γ specifies the guess rate and λ specifies a lapse rate, is used 1.1 [12, 13]. An example of a psychometric function with non-zero γ is shown in Fig. 1.2.

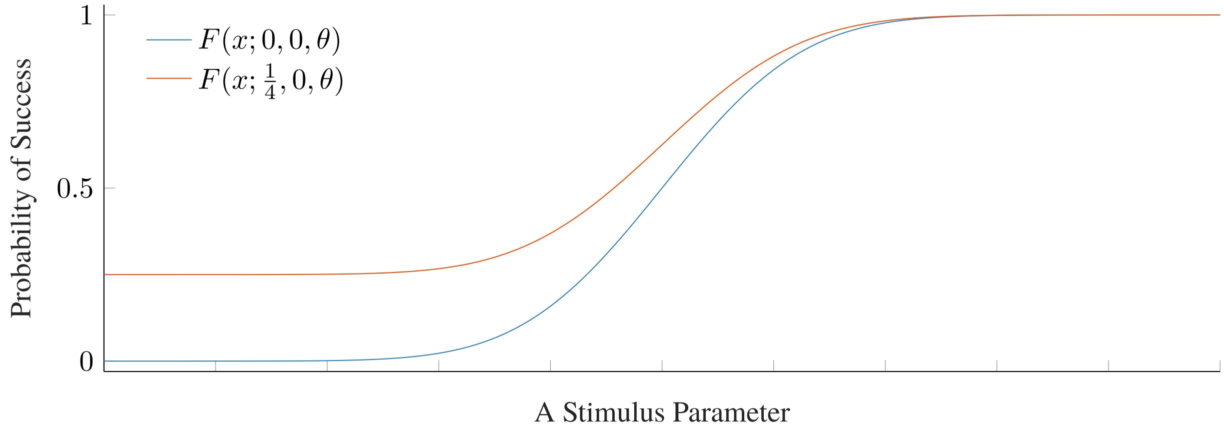

Figure 1.2: Psychometric Function with $\gamma = \frac{1}{4}$ compared to $\gamma = 0$

1.2 Fitting Psychometric Functions and The Gaussian Process

Traditionally, experiments to fit a psychometric function have involved repeatedly delivering tasks and making observations until a sufficient amount of data has been collected. If a random sampling scheme is used, this may require hundreds of observations and depending on the complexity of the task, this could easily be hours of subject testing time. This severely limits the kinds of tasks that can be delivered. Increasing the dimensionality of the stimulus further exacerbates this problem.

As mentioned previously, since a PF maps an input domain to a Bernoulli random variable, it may be used to form a likelihood function to determine its parameters given a set of task performance data $\mathcal{D}$.

$$\mathcal{L}(\theta \mid \mathcal{D}) = \mathcal{L}(\theta \mid x, y) = \begin{cases} \psi(x; \gamma, \lambda, \theta) & y = +1 \\ 1 - \psi(x; \gamma, \lambda, \theta) & y = -1 \end{cases} \tag{1.2}$$

If a specific functional form of sigmoid is selected, then its parameters can be directly estimated using maximum likelihood estimation or maximum *a posteriori* methods [12]. If the selected sigmoid is a logistic or probit function, then the problem reduces to that of logistic and probit regressions, respectively.

1.2.1 Gaussian Process Classification

In more recent works, non-parametric modeling of PFs has been done with Gaussian processes [10]. A Gaussian process (GP) is defined as "a collection of random variables, any finite number of which have a joint Gaussian distribution" [7]. In general, A GP over functions models the function value at each point in a domain as a Gaussian distribution such that the distribution of function values over an entire domain is the joint distribution of each individual Gaussian. A GP is fully specified by its mean $m(x)$ and covariance $k(x, x')$ functions and is often written in the form of Eq. 1.3.

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')) \tag{1.3}$$

Given a data set $\mathcal{D}$, a GP posterior distribution over function values is found by applying Bayesian inference to a prior distribution through a likelihood function. The relationship between these components is given by Eq. 1.4.

$$p(f \mid \mathcal{D}) \propto p(\mathcal{D} \mid f)p(f) \tag{1.4}$$

In Bayesian analysis, the prior distribution is typically thought to encode prior knowledge about the behavior of a random variable. However, a prior is not required to be informative as even an uninformative prior can be updated with data. The prior distribution for a GP model is also a GP that can encode information like the expected properties of a function (*e.g.*, linearity or smoothness). For this project, only uninformative priors are considered, though a GP resulting from a previous experiment or formed by averaging the results of several experiments can be extremely valuable as a starting point for new experiments, especially for optimizing the amount of data required.

The predictive distribution of a GP over the function values f_* for a set of test points x_* is given by

$$p(f_* \mid X, y, x_*) = \int p(f_* \mid X, x_*, f)p(f \mid X, y)\, df, \tag{1.5}$$

where $p(f \mid X, y)$ is the posterior distribution over function values resulting from Bayesian inference. For classification problems, a GP is placed over an unobserved, latent function. This latent function is analogous to the linear model that forms the basis of logistic regression. When making predictions, a predictive distribution (Eq. 1.5) over the latent function value is formed. Then, this distribution is then mapped to class probabilities π_* by computing the expected values of a sigmoid function $\sigma(\cdot)$ applied to the latent function as shown

$$\pi_* = \Pr(y_* = +1 \mid X, y, x_*) = \int \sigma(f_*) p(f_* \mid X, x_*, f) \, df. \tag{1.6}$$

This sigmoid function is usually the same as the likelihood function for classification.

The likelihood used in a GP must be Gaussian as this allows for a Gaussian posterior when using a Gaussian prior. This property of Gaussian distributions is referred to as self-conjugacy. In general, if a prior distribution and a posterior distribution are from the same family, they are said to be conjugate. However, since our likelihood function ψ is non-Gaussian, the posterior will not be Gaussian. Thus, we use numerical methods to compute an Gaussian approximation of the posterior distribution.

1.2.2 Hyperparameter Optimization

Although the Gaussian process model of a function's values is non-parametric, the component functions of a GP (*i.e.*, the mean, covariance, and likelihood functions) can all have parameters called hyperparameters. For example, we may specify that a GP has a prior mean that is a constant value across the entire domain. Such a mean is parameterized by θ and simply returns θ when evaluated.

$$m(x; \theta) = \theta \tag{1.7}$$

Finding the optimal value for θ is part of the training process of the model.

In Bayesian analysis, the normalization factor, or the marginal likelihood, is often referred to as the model evidence.

$$\text{marginal likelihood} = \log \int p(y \mid f)p(f \mid X)\, df \tag{1.8}$$

This is the integral of the likelihood times the prior over f and can be thought of as the total likelihood of the data given the latent function f. Since model evidence is a measure of fit, it can be used as an objective function for optimizing hyperparameters. In practice, this is often framed as minimizing the negative of the model evidence, which enables the use of existing convex optimization tools.

Chapter 2

A Generalized Likelihood Function

A likelihood function, $p(\mathcal{D} \mid f)$, imparts the ability to reason about some data to our model. As we saw in Chapter 1, we may use a generalized psychometric function (Eq. 1.1) as a likelihood function.

$$\Pr(y = +1 \mid x) = \psi(x; \gamma, \lambda, \theta) \tag{2.1}$$

The objective in this section is to describe the implementation of such a likelihood function for use with GPs, specifically with the GPML library [6].

2.1 ψ as a Mixture

Through a simple transformation, $\psi(\cdot)$ can be seen as a weighted mixture of a Bernoulli distribution $C(p)$ (further defined in Section 2.2) and a sigmoid likelihood function $F(x)$. The sigmoid $F(x)$ is the same as that used in the definition of ψ.

$$
\begin{aligned}
\psi(x;\gamma,\lambda,\theta) &= \gamma + (1-\gamma-\lambda)F(x;\theta) \\
&= (\gamma+\lambda)\frac{\gamma}{\gamma+\lambda} + (1-(\gamma+\lambda))F(x) \\
&= w\frac{\gamma}{\gamma+\lambda} + (1-w)F(x) \\
&= wC\left(\frac{\gamma}{\gamma+\lambda}\right) + (1-w)F(x;\theta)
\end{aligned} \tag{2.2}
$$

The GPML library provides functionality for implementing mixture likelihood functions. The implementation is general to any number of likelihood functions. For any two component likelihoods, denoted L_1 and L_2, it takes the form

$$M(x; w_1, w_2, L_1, L_2) = \log[w_1 L_1 + w_2 L_2], \tag{2.3}$$

where $w_1, w_2 \in [0, 1]$ are weights that sum to 1. For numerical stability, as the values produced by likelihood functions may be very small, Eq. 2.3 computes the log of the mixture likelihood. Thus, we can write $\log \psi$ as a function of M:

$$\log \psi(x;\gamma,\lambda,\theta) = M(\gamma+\lambda, 1-\gamma-\lambda, C(p), F(x;\theta)). \tag{2.4}$$

2.2 Implementing a Bernoulli Likelihood

In Eq. 2.2, a Bernoulli likelihood function, denoted $C(p)$, was used to provide a constant value p to a mixture likelihood. Such a likelihood function has the form

$$C(p) = \begin{cases} p & y = +1 \\ 1-p & y = -1 \end{cases} . \tag{2.5}$$

However, to allow for unconstrained hyperparameter optimization (see Section 1.2.2), the parameter p should be permitted to take any real value. This can be achieved by reparameterizing the function to be $C(\theta)$ such that the Bernoulli probability p is a sigmoid transformation of θ. During implementation, a Normal Cumulative Density Function (CDF), commonly denoted as Φ was used, though it should be noted that any continuous sigmoid function may be used (*e.g.*, a logistic function). This results in

$$C(\theta) = \begin{cases} \Phi(\theta) & y = +1 \\ 1 - \Phi(\theta) & y = -1 \end{cases}. \tag{2.6}$$

The implementation of this likelihood function for GPML requires considering two modes: prediction mode and inference mode. The first, prediction mode, is used to evaluate the predictive distribution for test points, the expected likelihood for those points over the posterior distribution,

$$q(y_*) = \int \mathcal{N}(f_*|\mu, \sigma^2)p(y_*|f_*)\ df_*. \tag{2.7}$$

In this equation, f_* represents the values of the eventually "squashed" latent function and $\mathcal{N}(f_*|\mu, \sigma^2)$ is the posterior distribution of those function values. Note that this is equivalent to Eq. 1.6.

In the case of the Bernoulli distribution, neither the probability of $y_* = +1$ nor its complement for the $y_* = -1$ case is dependent on f_*. Therefore, the predictive marginal distribution $q(y_*)$ is the same as the Bernoulli probability mass function (pmf),

$$q(y_*) = \begin{cases} C(\theta) & y = +1 \\ 1 - C(\theta) & y = -1 \end{cases}. \tag{2.8}$$

The first and second moments of are also the same as those for a Bernoulli distribution:

$$\mu_{y_*} = 2\beta - 1, \text{ and} \tag{2.9}$$

$$\sigma^2_{y_*} = 1 - (2\beta - 1)^2. \tag{2.10}$$

In inference mode, a likelihood function is used in conjunction with an inference method (e.g., Laplace Approximation) to compute a Gaussian approximation of the posterior distribution. This is because, as mentioned in Section 1.2.1, the Gaussian distribution is self-conjugate, which means that a Gaussian prior is only compatible with a Gaussian likelihood. The implementation of this Bernoulli likelihood developed for this project supports two inference methods: Laplace Approximation (LA) and Expectation Propagation (EP).

In addition to the likelihood function itself, Laplace Approximation requires the computation of the first, second, and third derivatives with respect to the latent function f as well. Again, for numerical stability, these are provided as the log likelihood and its derivatives. In this case, however, since the Bernoulli likelihood is not dependent on f, we find that the derivatives with respect to f are all 0 (Eqs. 2.11, 2.12, 2.13).

$$\frac{\partial}{\partial f} \log p(y \mid f) = 0 \tag{2.11}$$

$$\frac{\partial^2}{\partial f} \log p(y \mid f) = 0 \tag{2.12}$$

$$\frac{\partial^3}{\partial f} \log p(y \mid f) = 0 \tag{2.13}$$

Additionally, the log-derivatives with respect to both the latent function and the hyperparameter θ, must also be computed. This is complicated by the use of a sigmoid transformation that allows for unconstrained hyperparameter optimization. The derivatives are given by Eqs. 2.14, 2.15, and 2.16.

$$\frac{\partial}{\partial \theta} \log p(y \mid f) = \begin{cases} \dfrac{\Phi'(\theta)}{\Phi(\theta)} & y = +1 \\ -\dfrac{\Phi'(\theta)}{1 - \Phi(\theta)} & y = -1 \end{cases} \tag{2.14}$$

$$\frac{\partial^2}{\partial f \partial \theta} \log p(y \mid f) = 0 \tag{2.15}$$

$$\frac{\partial^3}{\partial f^2 \partial \theta} \log p(y \mid f) = 0 \tag{2.16}$$

For EP, the likelihood must compute the expected value of the likelihood function over latent function values, Z. Again, as the likelihood function does not depend on the latent function, this expected value is simply the likelihood function itself,

$$Z = \int C(\theta)\mathcal{N}(f \mid \mu, \sigma^2)\, df = C(\theta). \tag{2.17}$$

The derivatives of the log-likelihood with respect to μ and μ^2 are also required and are both zero. Additionally, the derivatives with respect to the hyperparameter θ are the same as those for LA.

2.3 Implementing ψ and Hyperparameter Optimization

As described previously, the formulation of ψ as a mixture likelihood function in Section 2.1 (Eq. 2.4) can be implemented for the GPML library using the provided mixture likelihood implementation. Additionally, the definitions of ψ in psychometric function estimation literature ([12], [13]) are agnostic to the choice of sigmoid function $F(x;\theta)$ used. Possible choices include Logistic, Gaussian, and Weibull CDFs, among others. In this work, a standard Normal CDF, denoted $\Phi(x)$, will be used for simplicity. Thus, for this choice of sigmoid, the ψ function can be rewritten once more as

$$\log \psi(x;\gamma,\lambda) = M(x;\theta_1,\theta_2,\theta_3) = \log[\theta_1 C(\theta_3) + \theta_2 \Phi(x)], \tag{2.18}$$

where $\theta_1 = \gamma + \lambda$, $\theta_2 = 1 - \gamma - \lambda$, and $\theta_3 = \frac{\gamma}{\gamma+\lambda}$. This notation reflects the choice in sigmoid function as well as the lack of parameters for that function.

As written, M can be implemented with GPML provided functions along with the custom Bernoulli likelihood described in Section 2.2. However, one caveat about this form is that the parameters of this function will be optimized directly. This means that γ and λ are only represented implicitly and cannot be interfaced with directly. While the variables can be solved for via a simple system of equations, there may be certain scenarios or experiments that require the arbitrary fixation of both variables to some theoretical values. It is necessary to reparameterize ψ one more time to enable this capability. As θ_1, θ_2, and θ_3 are functions of γ

and λ, it is trivial to programatically enforce the ψ interface instead of the M interface. The only special attention required is the application of the chain rule to the derivatives of M with respect to its hyperparameters to account for the reparameterization to γ and λ.

As in Section 2.2, to allow for unconstrained optimization, the parameters of this wrapper, denoted ψ', should be permitted to be any real value as $\gamma, \lambda \in [0,1]$. Thus, ψ' is parameterized by $g = \Phi^{-1}(\gamma), l = \Phi^{-1}(\lambda)$ and

$$\log \psi'(x; g, l) = M\left(\Phi(g) + \Phi(l), 1 - \Phi(g) - \Phi(l), \frac{\Phi(g)}{\Phi(g) + \Phi(l)}\right). \tag{2.19}$$

The partial derivatives of ψ' with respect to each of its hyperparameters can be computed by applying the chain rule to the gradient of M with respect to the hyperparameters of M. For example, the partial derivative of ψ' with respect to g is given by Eq. 2.20 with $\Theta = \{\theta_1, \theta_2, \theta_3\}$.

$$\frac{\partial}{\partial g}\psi'(x; g, l) = \sum_{\theta_i \in \Theta} \frac{\partial M}{\partial \theta_i}\frac{\partial \theta_i}{\partial g} \tag{2.20}$$

Chapter 3

Results and Conclusions

3.1 Validation Experiments

With a working implementation developed in Chapter 2, the next step is to demonstrate its results. All of the experiments presented below were completed using a constant mean function, a linear covariance function, and Laplace Approximation inference.

To begin, we can look at the performance of a GP model on psychometric data without using a generalized likelihood function (see Fig. 3.1). In this example, it can be seen that the model is unable to understand the non-zero guess rate for intensities less than 50 dB. Figure 3.2 shows the estimates made on the same observations by a model using a generalized ψ likelihood function. It is obvious that the estimates are improved as the model is able to account for the guesses and lapses of the observed data.

It is also possible to compare the GP estimate of the psychometric function to those of existing software. The Palamedes Toolbox ([5]) provides tools to form a maximum likelihood estimate of the ψ function parameters (see Eq. 1.1). Another library called Psignifit ([8]) can provide maximum *a posterior* estimates for the parameters. Figure 3.3 shows the estimates produced by these libraries along side the one produced by a GP using a generalized likelihood function. While the estimates for α, arguably the most significant result of modeling a psychometric function, are consistent between all three estimates, the estimate for β is not.

Furthermore, it appears that the GP model is more robust to noise in the data, producing a similar model across all three data set sizes.

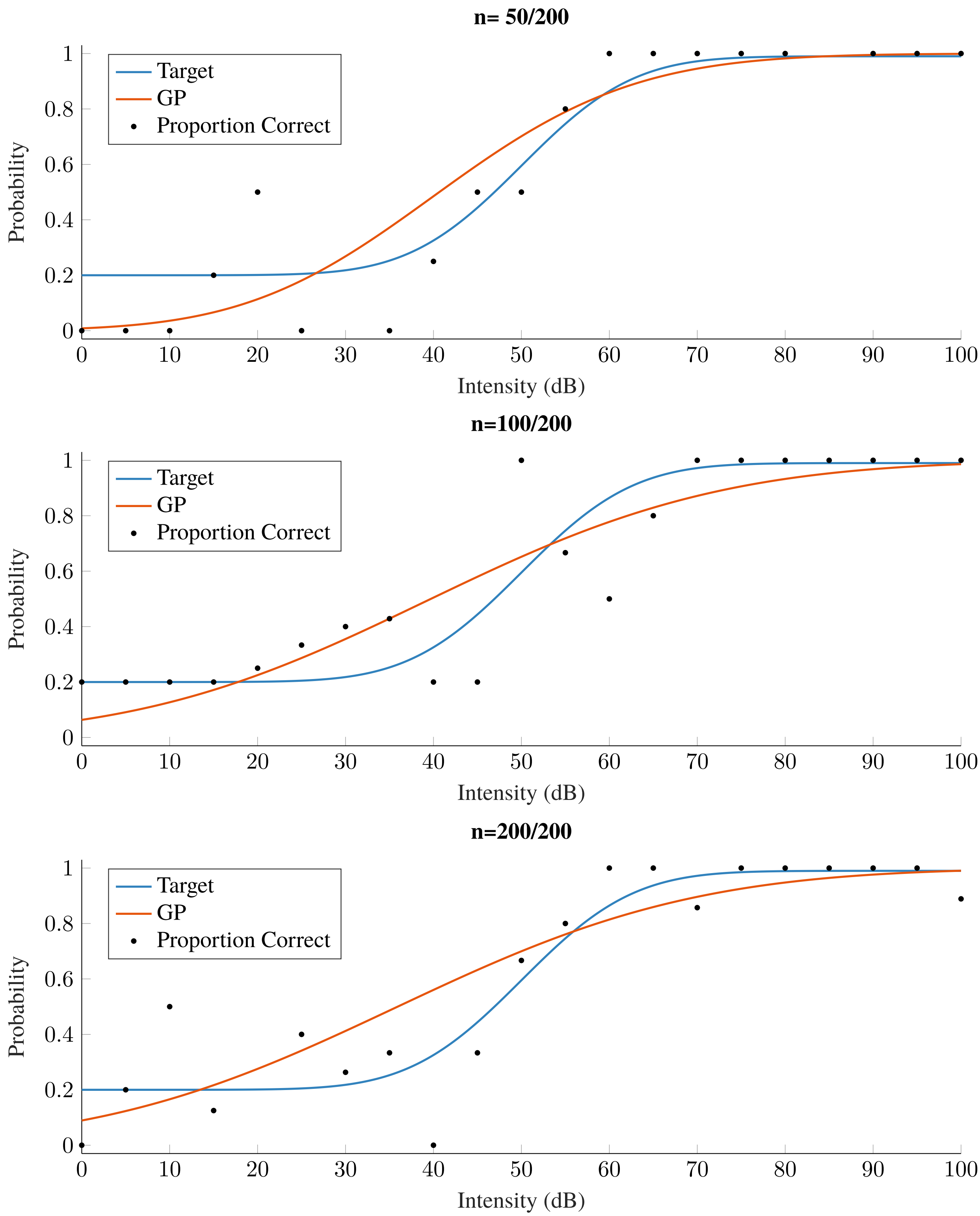

Figure 3.1: Baseline Model Performance

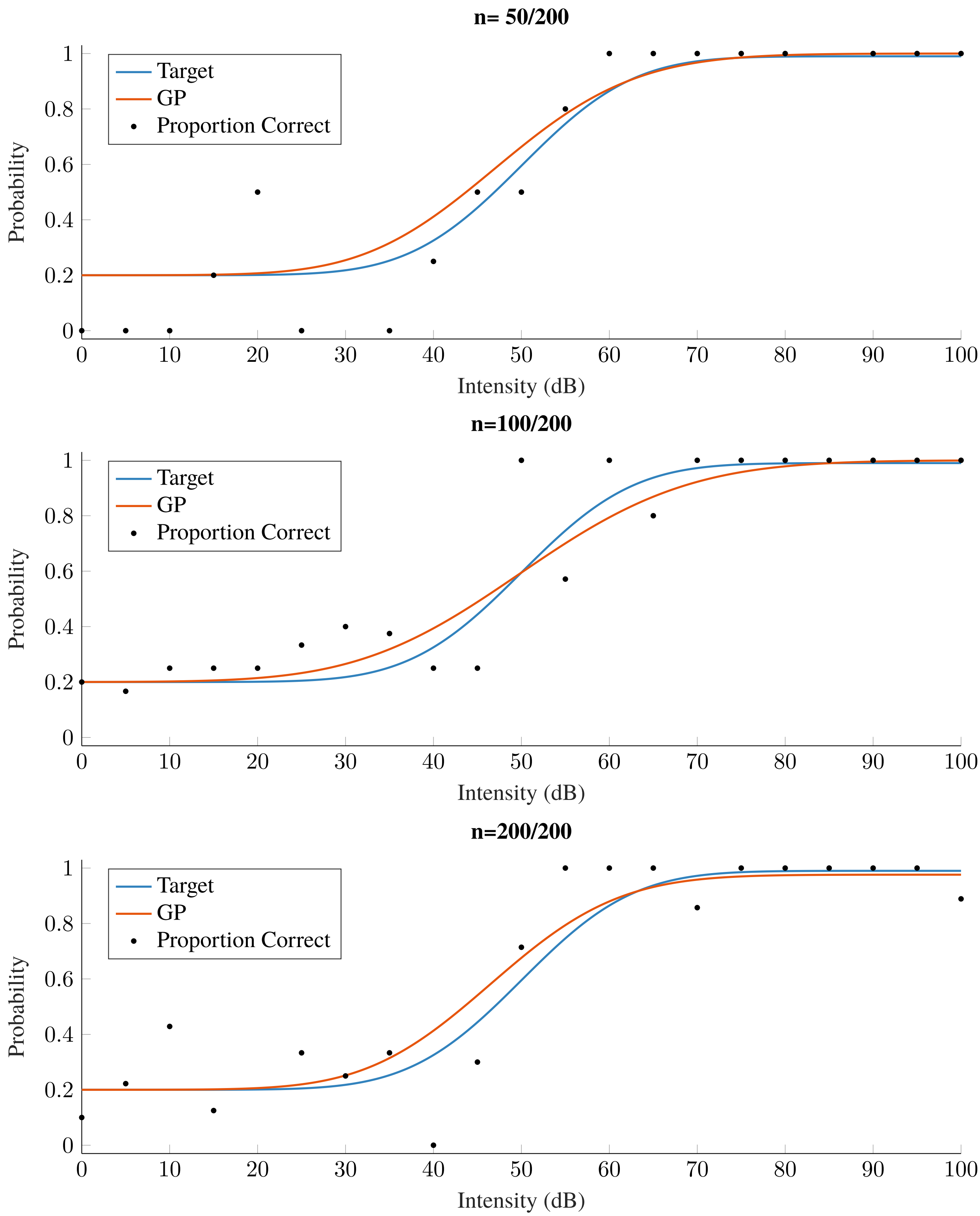

Figure 3.2: Model Performance with Generalized Likelihood Function

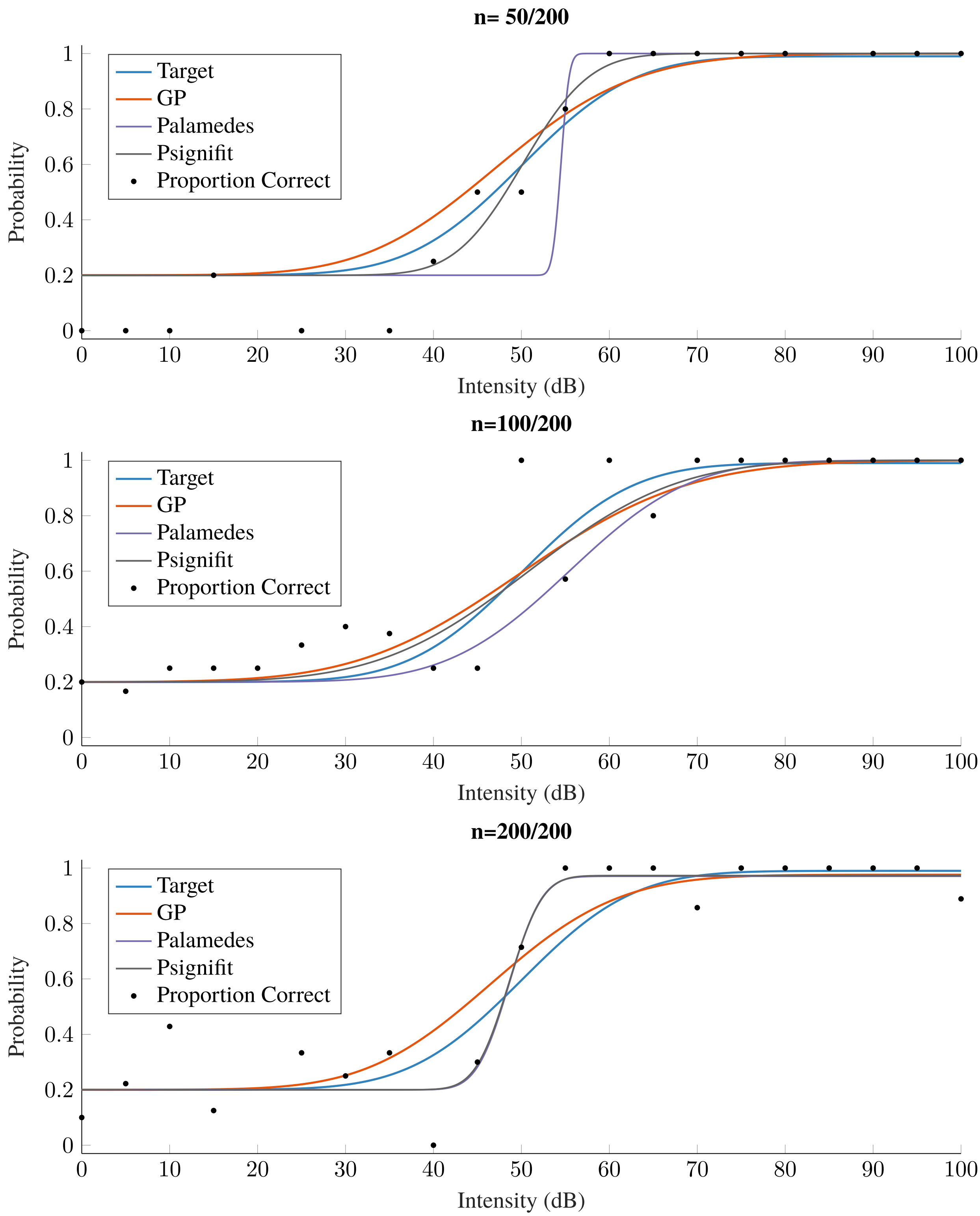

Figure 3.3: Palamedes and Psignifit Estimates of Psychometric Functions

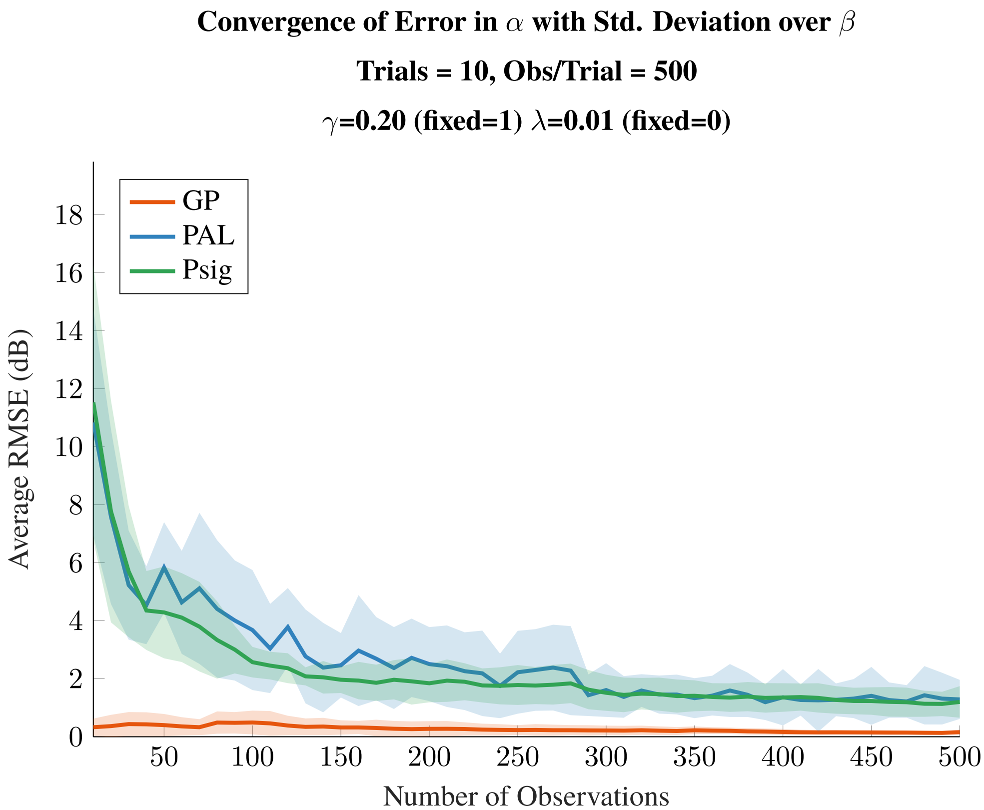

Figure 3.4: Average RMSE of α Across All β.

To further analyze the performance, the average root mean squared error (RMSE) between predicted values and true values for α and β were computed across 7 different β values (0.2, 0.5, 1, 2, 5, 10, 20). Ten trials for each β, each with estimates on data sets of increasing size up to 500 observations in steps of 10 were completed. The results are shown in Figures 3.4 and 3.5. From these figures, it can be seen that the GP estimate of both α and β parameters are consistently better than those of Palamedes or Psignifit.

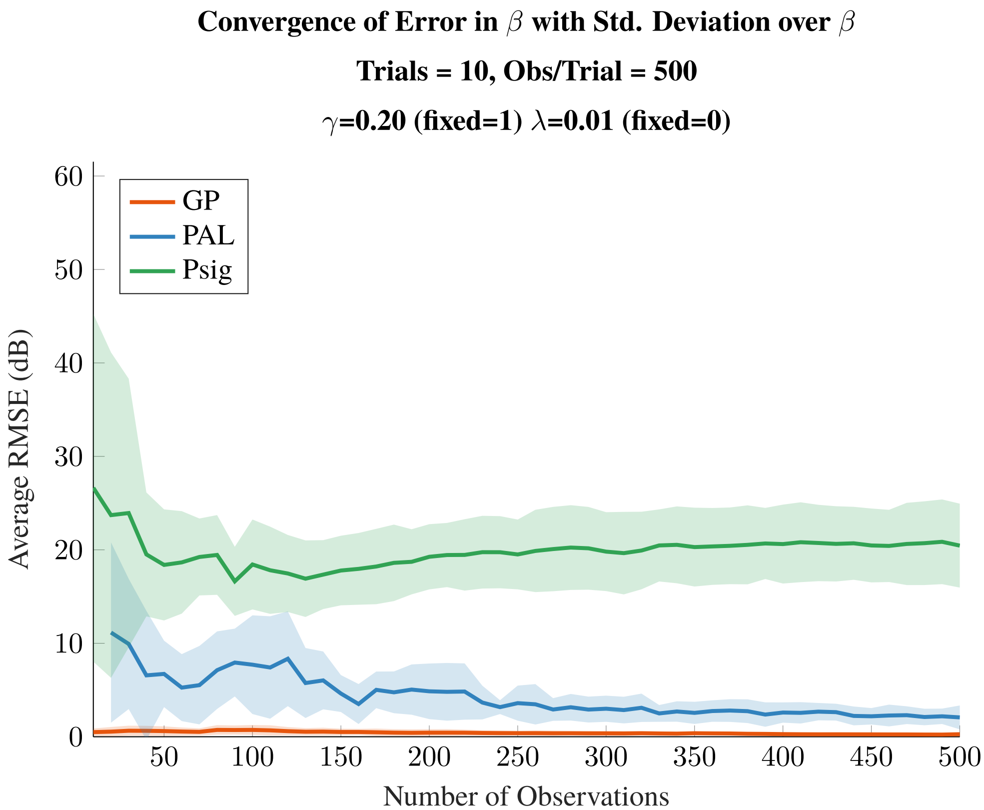

Figure 3.5: Average RMSE of β Across All β.

3.2 Next Steps

A generalized psychometric function likelihood for Gaussian Processes enables the modeling of complex perceptual and cognitive phenomena represented by tasks with theoretical guess and lapse rates. This is true in terms of both fitting a model and improving experimental efficiency.

A Gaussian Process model allows for the simple modeling of higher dimensional psychometric functions, where only a subset of the dimensions scale linearly in probability. For example, in hearing, although the probability that a subject can hear a tone increases monotonically as intensity increases, this is not the case in the frequency dimension. An example of such a psychometric function is shown in Fig. 3.6. A GP can be extended to model such psychometric functions with little additional configurations. This is not true for

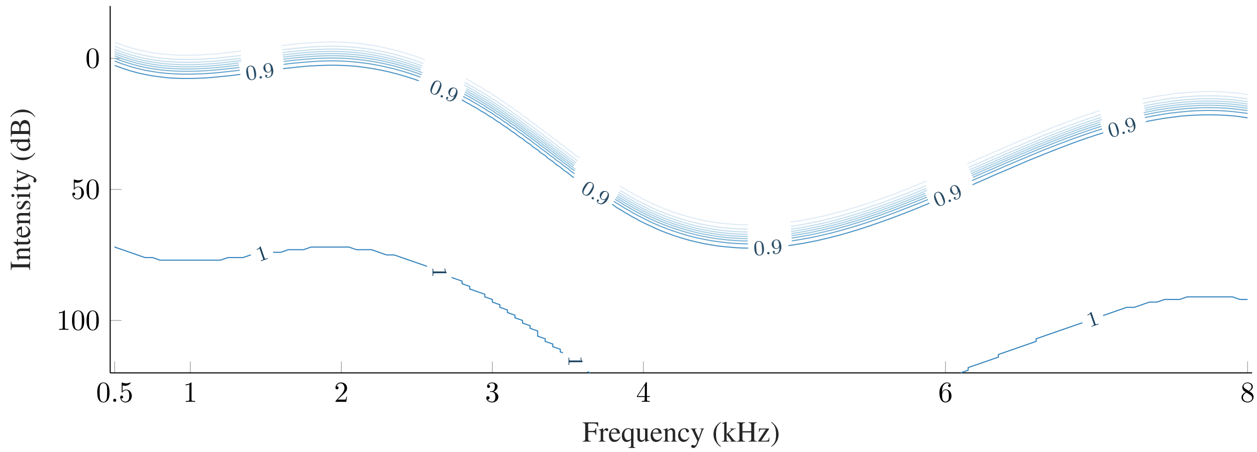

Figure 3.6: An Example of a Two Dimensional Hearing Test (Audiogram)

parametric methods such as MLE, as one would be required to specify an underlying model form (*e.g.*, a linear model).

Additionally, the posterior GP produced using a generalized likelihood function also enables the application of active learning to these tasks [11]. Active learning takes advantage of the posterior distribution to reason about where the model is most uncertain. Making observations at these points have high potential to produce the most information gained. If each observation is made optimally, the total number of observations required to fit a psychometric function may be reduced. This shortens the amount of time required for a subject to complete a task, greatly improving the accessibility and feasibility of many psychometric tasks.

Further investigation and development in both of these areas are likely to bring breakthroughs in psychometrics, as well as in other areas, such as education and medicine. The development of a generalized likelihood function for psychometric function estimation is only one step in these directions.

References

[1] Edwin V Bonilla, Kian M. Chai, and Christopher Williams. Multi-task Gaussian Process Prediction. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 153–160. Curran Associates, Inc., 2008.

[2] H. Levitt. Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B):467–477, February 1971.

[3] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with Noisy Labels. page 9.

[4] Emmanuel Ponsot, Léo Varnet, Nicolas Wallaert, Elza Daoud, Shihab A. Shamma, Christian Lorenzi, and Peter Neri. Mechanisms of spectrotemporal modulation detection for normal- and hearing-impaired listeners. *bioRxiv*, page 2020.01.03.894667, January 2020.

[5] Nicolaas Prins and Frederick A. A. Kingdom. Applying the Model-Comparison Approach to Test Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox. *Frontiers in Psychology*, 9, 2018.

[6] Carl Edward Rasmussen and Hannes Nickisch. Gaussian Processes for Machine Learning (GPML) Toolbox, December 2010.

[7] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, Mass, 2006. OCLC: ocm61285753.

[8] 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, May 2016.

[9] Clayton Scott, Gilles Blanchard, and Gregory Handy. Classification with Asymmetric Label Noise: Consistency and Maximal Denoising. In *Conference on Learning Theory*, pages 489–511, June 2013.

[10] Xinyu D. Song, Roman Garnett, and Dennis L. Barbour. Psychometric function estimation by probabilistic classification. *The Journal of the Acoustical Society of America*, 141(4):2513, April 2017.

[11] Xinyu D. Song, Kiron A. Sukesan, and Dennis L. Barbour. Bayesian active probabilistic classification for psychometric field estimation. *Attention, Perception & Psychophysics*, 80(3):798–812, April 2018.

[12] Bernhard Treutwein and Hans Strasburger. Fitting the psychometric function. *Perception & Psychophysics*, 61(1):87–106, January 1999.

[13] F. A. Wichmann and N. J. Hill. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8):1293–1313, November 2001.