

Article

Prior Distribution and Entropy in Computer Adaptive Testing Ability Estimation through MAP or EAP

Joel Suárez-Cansino ¹, Virgilio López-Morales ^{1,*}, Luis Roberto Morales-Manilla ²,
Adrián Alberto-Rodríguez ¹ and Julio César Ramos-Fernández ³

¹ Basic Sciences and Engineering Institute, Systems and Information Technologies Research Center, Intelligent Computing Research Group, Autonomous University of Hidalgo State, Col. Carboneras, Mineral de la Reforma 42184, Hidalgo, Mexico

² Department of Software Development, Advanced Computing and Innovation Research Group, Polytechnic University of Tulancingo, Tulancingo 43629, Hidalgo, Mexico

³ Department of Mechatronics, Smart Technologies Applied to Social Development Research Group, Polytechnic University of Pachuca, Zempoala 43830, Hidalgo, Mexico

* Correspondence: virgilio@uaeh.edu.mx

Abstract: To derive a latent trait (for instance *ability*) in a computer adaptive testing (CAT) framework, the obtained results from a model must have a direct relationship to the examinees' response to a set of items presented. The set of items is previously calibrated to decide which item to present to the examinee in the next evaluation question. Some useful models are more naturally based on conditional probability in order to involve previously obtained hits/misses. In this paper, we integrate an experimental part, obtaining the information related to the examinee's academic performance, with a theoretical contribution of maximum entropy. Some academic performance index functions are built to support the experimental part and then explain under what conditions one can use constrained prior distributions. Additionally, we highlight that heuristic prior distributions might not properly work in all likely cases, and when to use personalized prior distributions instead. Finally, the inclusion of the performance index functions, arising from current experimental studies and historical records, are integrated into a theoretical part based on entropy maximization and its relationship with a CAT process.

Keywords: entropy; CAT; Kullback–Leibler divergence; likelihood function; maximum a posteriori; expectation a posteriori; Bayesian inference; performance index function; item response theory; item characteristic curve



Citation: Suárez-Cansino, J.; López-Morales, V.; Morales-Manilla, L.R.; Alberto-Rodríguez, A.; Ramos-Fernández, J.C. Prior Distribution and Entropy in CAT Ability Estimation through MAP or EAP. *Entropy* **2023**, *25*, 50. <https://doi.org/10.3390/e25010050>

Academic Editor: Antonio M. Scarfone

Received: 20 September 2022

Revised: 29 November 2022

Accepted: 14 December 2022

Published: 27 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

When one wants to explain some relationship between latent traits of individuals, for instance, unobservable characteristics or attributes and their manifestations (observed outcomes, responses, or performance), then item response theory (IRT) becomes a valuable formal tool. IRT is a family of models to analyze and predict the behavior of the involved variables, and their applications cover different assessment scenarios. Item exposure control, item calibration, and automatic item generation are only some examples of these scenarios and the involved variables. From a theoretical point of view, there are several well-known models to analyze these topics and potentially helpful tools to propose novelty solutions [1–5]. These models consider a set of items to define a measurement instrument. One set of parameters specifies the item's characteristics that depend on the particular application. The results of the application of the measurement instrument provide information about the examinee's latent trait. IRT assumes that the latent construct values (e.g., stress, knowledge, attitudes values) and some items' parameter values are organized in an unobservable continuum as random variables. Thus, IRT helps to establish

the position or value of the examinee's latent trait on that continuum by considering the items' characteristics and the quality of responses to them [6,7].

Items with different presentation and answer formats, examinees, and examiners are just a part of the assessment scenario. Particularly, computer adaptive testing (CAT) is an area where the application of IRT is highly useful to automatize performance assessments.

In general, IRT and its application to CAT assume the existence of a pool of items that by construction has finite cardinality. The CAT process supposes that, through experimentation, the pool contains calibrated items (calibrated items pool or CIP), i.e., a previous experiment provides information about the values of the items' parameters that define its corresponding characteristics.

The item calibration process entails a statistical analysis of the responses arising from a set of test subjects, and the fitting of the experimental data to a sigmoidal cumulative distribution function (CDF) model assigned to every item. This procedure defines the corresponding item's parameters and, therefore, the item characteristic curve (ICC) which depends on the latent trait as *ability*. The parameters of an ICC have an interpretation in terms of the item's difficulty and the item's discrimination capability, among others, and they influence the determination of the value of the latent trait under analysis. There are some well-known ICC models, such as 1PL, 2PL, and 3PL, although one can find 4PL and 5PL models, too. Naturally, 1PL (one-parameter logistic model) and 2PL (two-parameter logistic model) are the simplest models since their parameters have a direct interpretation and their relationships with the process of searching for the latent trait are clear.

Equation (1) defines the general structure for the 2PL model and gives the conditional probability of correctly answering an item with known difficulty and discriminant:

$$P(\theta|\mu, \alpha) = \frac{1}{1 + e^{-\alpha(\theta-\mu)}}, \quad (1)$$

where the parameters μ and α represent, respectively, the item's difficulty and the item's discriminant, and θ refers to the examinee's ability.

An examinee participating in a CAT has to pass through three steps.

1. Assignment of an initial estimate of the examinee's ability or the item's difficulty, since the system needs to know the characteristics of the first item in the evaluation process.
2. The system saves the examinee's response, decides if the examinee gave a correct answer or not, and builds the response pattern for this specific testing process.
3. The system considers the response pattern and the selected latent trait model to build a likelihood function, intending to decide what proper item (calibrated) must come next. There are several methods to do this, and here we apply the definition of the prior distribution. After deciding what item comes next, the CAT procedure poses this item to the examinee, and the testing returns to the second step again. Our main contribution aims to solve some problems in this step.

A reliable estimate of the next question to be presented, for instance, through the concept of the maximum likelihood function, requires at least two responses to the presented items in the evaluation process. One of the items needs to have a right answer, and the other an incorrect one. Only in this case, the likelihood function will have an extreme point in the set of values of abilities and, therefore, a maximum value at this point.

Note that in the event that all the answers obtained were correct (or incorrect), the likelihood function is just a sigmoid (the ICC of the items) that does not have extreme points in the domain of the examinee's ability. Because of this reason, it becomes impossible to compute the next estimation of the examinee's ability.

To overcome the difficulty of estimating the next examinee's ability by just using the likelihood function, some authors have proposed different options:

1. The use of two fictitious items with high and low probabilities to ensure that the examinee answers alternatively correctly and incorrectly to the items.

2. The use of heuristic formulas to estimate the examinee's latent trait until a maximum likelihood makes it possible to estimate the value.
3. To define prior distributions until one can apply a likelihood function to estimate the examinee's ability. This proposal relates directly to the statement of the research problem in this paper.

The first option has the inconvenience that the estimated latent trait value after applying the first non-fictitious item reaches very extreme values, and the second non-fictitious item of the CAT process provides more information for that extreme value of ability. Thus, the second non-fictitious item becomes less informative for the final ability value and it does not contribute considerably to the test precision [8,9].

The second option has the inconvenience that in some circumstances the CAT process does not converge, although when the increment (or decrement) of the latent trait value is variable this phenomenon does not occur [8,9].

Finally, the third option has several inconveniences:

- (i) A general use of prior information in educational assessment appears to be inhibited solely by the assumption that including a priori information on test scores in performance assessment may be unfair to students [10].
- (ii) Assuming that regular evaluation practices include information provided by the examinee (regarding past experiences) or collected from multiple sources in the assessment procedure without specifying the type of the sources [10].

Additionally, there is not much information about the potential risks when prior information is not perfectly accurate. Overconfidence in inaccurate prior information may in fact increase test length and/or lead to severely biased final latent trait estimates. In this event, then the system could, for example, select an incorrect starting point or introduce bias in the trait estimation process, and provide items that do not match the participant's trait level [10]. On the other hand, the level trait does not depend solely on the examinee's performance but on the values of mean and variance that one assigns to the trait's prior distribution in the population [8].

From a theoretical point of view, depending on the established a priori distribution, one can obtain a multimodal posterior so that the Bayesian MAP estimation might refer to a local maximum [8].

Finally, in some cases, the Bayesian procedures provide estimation with a specific regression toward the mean of the prior distribution of the latent trait. This phenomenon can favor examinees with low levels and affect examinees with high ability [8].

There are several advantages and drawbacks of introducing information before starting an adaptable evaluation process. The usual way of building prior distributions lends itself to subjectivities, even though the benefits in the administration of the evaluation are undoubted [11,12]. However, the subjectivity inherent in the prior distribution can be minimized as long as reasonable evidence supports the distribution proposal [11].

In this work, we address the role that the entropy can play in the reduction of this subjectivity in the construction of the prior distribution by using a set of proposed constraints related to entropy. Any distribution must satisfy these constraints that consider, for instance, its first and second moments and the academic framework such as, for example, school dropout and failure, among others.

1.1. Problem Statement

The use of Bayesian statistical inference in the CAT process is delicate and has to justify the application of essential components such as the prior distribution [13]. Different theoretical and experimental techniques exist to determine the prior distribution to initialize the CAT process. Some authors suggest that physical, mathematical, engineering, expert opinion models, historical data under similar circumstances, or other reasonable information can support the prior proposal [13]. Thus, we formally introduce the models related to academic performance, for example, the failure rate, the dropout rate, the study

habits index, and the subject comprehension index, among others, to further specify the structure of the prior distribution by using the concept of entropy.

1.1.1. Preliminaries

The estimation of the ability of a test subject presents problems at the beginning of the evaluation process when using the maximum likelihood method and when the examinee responds correctly or incorrectly to all the test items. Several proposals solving this problem have been published and there are some options based on Bayesian inference [14]. In particular, the MAP or EAP techniques use the concept of the prior distribution, with the drawback that the definition of the structure of this distribution can lead to subjectivities.

1.1.2. Originality

Within the given context, there is not enough information about the best prior distribution to be selected. Due to the Bayesian nature, MAP or EAP techniques require previous knowledge of the prior distribution, which contains initial statistical information about the ability of the examined subject.

One typically uses a normal distribution [15–17], but there is no evidence that this is necessarily correct since there is no reliable way to support the decision to opt for one prior type of distribution over another. The initial choice of the a priori distribution is paramount since it directly affects the calculation of the skill estimate and other parameters.

Furthermore, the structure of the psychometric model supporting the Bayesian inference process must be considered. An adequate structure selection provides an appropriate interpretation of each item's characteristics, predicts the consequences of using a psychometric model with the selected characteristics, and ensures the relationship between these options and the multimodality and bias characteristics in the a posteriori distribution that finally helps to estimate the corresponding latent trait [18–21].

1.1.3. Impact

In order to solve the former problems, one must then propose the form of the prior distribution through formal criteria to select good prior distributions. Some authors define some non-formal criteria and give quite illustrative examples of how the selection of a priori distribution affects the posterior distribution [22,23]. However, this research paper works mainly with the concept of entropy and, in a first instance, with the definition proposed by Shannon [24].

1.2. Article Structure

The paper is organized as follows: Section 2 focuses on a short hypothesis or conjecture statement and the paper's objectives. Section 3 contains a brief discussion about some works on the importance of the prior distribution and the most common assumptions that the researchers make on its structure. This part also discusses the role that entropy could play in determining the a priori distribution and the previous work in this regard, but not within the framework of a CAT.

Section 4 briefly describes the differences between the 1PL and 2PL latent trait models and explains the meanings of the difficulty and discriminant parameters. Through these models and definitions, the concepts of maximization a posteriori, or MAP, and expectation a posteriori, or EAP, and their relationships with the prior distribution are introduced.

In addition, one recalls Shannon's concept of entropy and states the ansatz (assumptions about the form of an unknown function, made to facilitate the solution of a problem) that give rise to the proposed method for estimating the prior distribution. Section 5 illustrates our numerical experimentation results, and provides and discusses the findings about the structures of the a priori distributions obtained through the proposed method. Finally, Section 6 synthesizes the results from numerical experiments and remarks some comments about the future work within the topic of the paper.

2. Hypothesis or Conjecture Statement

The specification of the prior distribution is a problem that does not have a straightforward solution in a CAT process. Part of this is due to the lack of formal procedures to get an analytical form of the distribution since there is no standard procedure on how the required information, to start the CAT process, can be integrated into a methodology to get an approximation to the model defining acceptable prior distributions.

2.1. Hypothesis Statement

If there is no formal procedure to determine prior distributions to initialize the CAT process, and Shannon's entropy plays the role of the objective function depending on the *a priori* information distribution, which is subject to constraints of normality, mean values, and variance of the ability, in addition to the satisfaction of academic performance constraints considering failure, study habits, subject comprehension, and dropout rates of the course of interest, among others, then the formal finding of a prior distribution to initialize a CAT process is possible.

2.2. Objectives

Our general objective is to build informative prior distribution functions by considering the maximization of Shannon's entropy as a cost function that depends on the distribution of *a priori* information, subject to normality, mean, variance, and academic performance constraints, to obtain formal prior distribution expressions. The specific objectives are the following:

1. To propose an ansatz about the school performance of the examinees, considering that they must be random functions depending on the random variable defined by the latent trait θ and some specific parameters, through the analysis of qualitative results obtained by various authors and, with these results, subsequently introduce distribution constraints based on the proposed assumptions.
2. To build an objective function to maximize entropy by considering the definition of entropy and the proposed ansatz in objective 1, and to obtain a methodology building and applying prior distributions in the CAT process.
3. To obtain experimental results numerically by simulating the behavior of the CAT process, to later make comparisons of the advantages and disadvantages of different scenarios that use prior distribution estimations.

3. State of the Art

The determination of the *a priori* distribution is experimental or through consultation with experts. However, regarding the role that entropy can play in searching for an adequate prior distribution, one can find a few research works on the topic. In this sense, to know how the *a priori* distribution behaves, one needs prior knowledge of the properties that it may have (the normality of the distribution is the simplest example of this knowledge, but there may be some other properties that are possible to know beforehand) [25,26].

The concept of prior distribution plays a fundamental role in Bayesian inference, so experimental determination of how to obtain these distributions and what theoretical methods should be to get something similar are paramount. To build a prior distribution, it is first necessary to specify representative random variables. In this sense, there are several possibilities that this paper introduces.

In the first instance, one assumes that the prior distributions must be related to the parameters of the selected psychometric model and the examinee's latent trait variable to evaluate as proposed in [27–30] through the experimental construction of the corresponding prior distributions.

Additionally, one can consult experts in the knowledge domain to evaluate in order to obtain an opinion about the form or structure that the *a priori* distribution should have [28,30].

Despite not being connected to CAT systems, one can find in the literature some theoretical attempts to determine the structure of the prior distribution using the concept of entropy [26]. In addition to possibly getting the expert's opinion, no known procedure integrates the results of the experimental process which, with a theoretical basis, can specify the characteristics or conditions under which one can obtain adequate prior distributions; that is, leading to unbiased posterior distributions, without multimodality, and to reliable latent trait estimates [18–21].

From a theoretical point of view, some contributions have dealt with the topic of informative and non-informative prior distributions, and they apply these definitions as academic examples to show the effects that the *a priori* distribution has over the *a posteriori* distribution [31].

In practice, heuristic distribution applications are analyzed when they are not supported by experimental data. In fact, some authors state that the practical consequences of using a prior distribution can depend on data. A heuristic distribution, such as the uniform or the normal with zero mean and unitary variance, can lead to nonsense inferences even when it has a large sample size. Currently, the study of prior distributions becomes relevant to analyze problems inside the frontiers of applied statistics [32,33].

In this sense, our paper integrates the experimental part of obtaining information related to the examinee's academic performance into the theory of maximum entropy. The structure of the academic performance index functions supports this experimental part, which, as an additional result, explains under what conditions one can use heuristic priors. Additionally, the paper remarks that the heuristic prior distributions could not properly work in all the cases and that one must consider personalized prior distributions instead. Finally, the inclusion of the performance index functions, arising from current experimental studies and historical records, are integrated into a theoretical part based on the entropy maximization and its relationship with a CAT process.

4. Modeling Initialization of the Evaluation Process

Geometrically, the characteristic curve of an item with difficulty μ_1 differs from another associated with an item with difficulty μ_2 by a simple shift to the left or right on the domain of the ICC, which is given by the latent trait values, depending on whether $\mu_1 < \mu_2$ or $\mu_2 < \mu_1$, respectively, as shown in Figure 1. On the other hand, the 2PL latent trait model has a correspondence rule given by Equation (1), where the parameter α represents the discriminatory capacity of the item; that is, how well it differentiates between examinees who have a latent trait greater than difficulty μ and those who have an ability less than μ . In this case, the graphs of two items differ not only by the displacement produced by the difficulty parameter μ but also by the function's increasing rate, which is proportional to the parameter α (see Figure 1).

After the CAT system poses the first item to the examinee and obtains the answer, then the next step selects a second item within the CIP with characteristics depending on whether the answer is correct or incorrect. After answering the second item, the system selects the third item depending on the answers given to the first two items, whose response's configurations are in the following set:

$$\{(correct, correct), (correct, incorrect), (incorrect, correct), (incorrect, incorrect)\},$$

and so on until the evaluation process of an examinee ends.

When an examinee has answered n items, the number of configurations of n Bernoulli-like trials that are elements in the derived set is 2^n . In each case, a *sui generis* trajectory leads to the estimated value of the latent trait θ associated with the specific examinee. Naturally, in this case, the items are dichotomous.

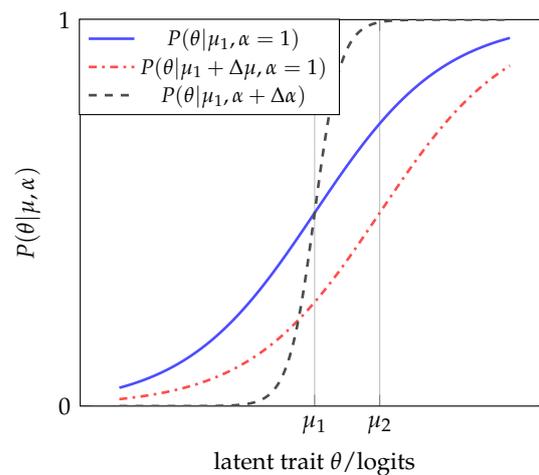


Figure 1. Effect of the difficulty value μ and the discriminant value α of an ICC in the case of the 1PL and 2PL latent trait models.

One of the main characteristics of a CAT is that the test should have the smallest possible number of items and still estimate the value of the specific ability, i.e., the selection of the n items in a particular sequence is not arbitrary. Given the response sequence for the first $n - 1$ items, it is possible to estimate the $(n - 1)$ -th value of the latent trait θ , which has the symbolic representation θ_{n-1} .

By knowing this estimate of the latent trait at iteration $(n - 1)$ -th, the CIP provides the next most informative item [8,34,35]. Fortunately, the Fisher’s information index gives a criterion to select the most informative one (see Equation (2)),

$$I(\theta) = \frac{\left(\frac{d}{d\theta}P(\theta|\vec{p})\right)^2}{P(\theta|\vec{p})Q(\theta|\vec{p})}, \tag{2}$$

where \vec{p} is the vector of parameters defining the structure of the latent trait model correspondence rule and $Q(\theta|\vec{p}) = 1 - P(\theta|\vec{p})$. For the 1PL and 2PL models, $I(\theta)$ is given by Equations (3) and (4), respectively.

$$I(\theta) = P(\theta|\mu)Q(\theta|\mu), \tag{3}$$

$$I(\theta) = \alpha P(\theta|\mu, \alpha)Q(\theta|\mu, \alpha). \tag{4}$$

Under the condition of independence and identical distribution of the items in the CIP, it is possible to build a likelihood function with the first $(n - 1)$ ICCs that the CAT system has applied up to the current answered items. In the best case, this likelihood function will have extreme points in the domain given by the latent feature values, implying that the likelihood function has at least one maximum [20,36].

However, the worst-case scenario is that all the first $(n - 1)$ items have a correct answer or all have a wrong answer. If one of these situations occurs, then building a likelihood function with a maximum, at least, is impossible. How does one determine the estimate of the latent trait value, in this case, to continue with the adaptive testing process?

There are several solution proposals to this problem, but a natural one [37] involves statistical information before the start of the evaluation process by using a Bayesian procedure. The idea is to use a prior distribution with which it is possible to use Bayesian argumentation to obtain estimates of the latent trait. Algorithm 1 provides a simple outline of this process.

Algorithm 1 Outline of the computerized adaptive assessment process.

```

1: procedure EVALUATIONPROCESS
2:    $item_1 \leftarrow$  select the first item with parameters  $\vec{p}_1$ 
3:    $top: response \leftarrow$  reply to item  $i - th, 1 \leq i$ 
4:    $\theta \leftarrow$  estimate latent trait based on the pattern of first responses to the  $i$  items
5:   if pattern is all correct or incorrect then
6:     return use prior distribution and Bayes
7:   else
8:     return use Maximum Likelihood estimation
9:    $item_{i+1} \leftarrow$  select item with a higher Fisher information in the Items Pool
10:  goto top

```

Note that the selection of the first item in step 2 of Algorithm 1 can proceed in at least one of two possible manners, namely

1. To calculate an estimate of the latent trait θ before starting the evaluation process and, with this estimate, to determine the item with the maximum Fisher information within the CIP [37].
2. To compute an estimate of the parameters of the first item (difficulty, discriminant, guessing, etc.) following some of the methods in [37].

Step 5 is central to the Algorithm 1 since Bayes' Theorem requires a *prior* distribution. Bayes' theorem involves the use of a prior distribution to calculate the so-called posterior distribution. However, selecting an *a priori* distribution is not trivial, and one must ensure that this distribution provides the highest amount of information about each of the examinees.

The following steps are essential to the understanding of our methodology:

1. To know the relationships among the *a posteriori* probabilities, the prior probability and the likelihood function.
2. Find the *a priori* probability and its closest dependence on an academic framework.
3. The analysis of discrete and continuous cases (the latter being of greater interest).

Regarding the first step, and given in Equation (5),

$$p(\theta|\vec{p}) = \frac{L(\vec{p}|\theta)}{p(\vec{p})} \cdot p(\theta), \quad (5)$$

we note that the likelihood function $L(\vec{p}|\theta)$ is the product of the item characteristic curves that arise throughout a specific individual evaluation pattern result. In this case, $p(\theta)$ directly gives the prior distribution.

Thus, the prior distribution is a function of the latent ability or trait θ . Finally, the transition from the discrete case to the continuous one is provided by:

$$\mathcal{S}(p) = - \int_{-\infty}^{+\infty} p(x) \log p(x) dx, \quad (6)$$

which may be subject to constraints of the form

$$\int_{-\infty}^{+\infty} f(x)p(x)dx = \langle f(X) \rangle, \quad (7)$$

where $\langle f(X) \rangle$ is the expectation of the random variable defined as $f(X)$, where X is a random variable whose values x 's define the population of interest. Equation (7) provides the general form of the constraints.

At this stage of the CAT process, the prior distribution $p(\theta)$ and the likelihood function $L(\vec{p}|\theta)$ are available to compute the posterior distribution $p(\theta|\vec{u})$ through Equation (5).

Once one calculates the posterior distribution, then estimates the next latent trait value, there are two possibilities:

- Determine the extreme point in the domain of the posterior distribution and compute the maximum value of the distribution at this extreme value.
- Determine the mean value of the latent trait population along the whole domain of the posterior distribution.

In order to sketch how the informative prior distribution can be related to the academic framework of the examinees, we propose several ansatzes.

Ansatz for Different Indices of Student Performance as a Function of the Ability θ

Some works use the concept of entropy [38] as an alternative for the construction of informative prior distributions. In this paper we introduce the maximum entropy through the application of optimization techniques to maximize the information that the entropy will yield concerning the specific examinee.

In addition to the distribution normality constraint, the latent trait mean and variance specifications, we analyze the contribution of special examinees' academic performance constraints to properly determine the population distribution through entropy maximization. In this sense, we apply the concept of an index (a random variable), depending on the ability θ .

By defining entropy as a cost function, entropy maximization considers that this function is subject to a list of constraints other than the constraints based on normality and first and second moments. The additional elements of the list of restrictions include the dropout rate, the failure rate, and the habits of study rates from one or more courses belonging to an examinee's record. Additionally, one can consider the index of understanding of topics that an examinee has in a historical academic record.

To relate study habits rate and its relationship to an ability function, several authors [39–43] have identified some factors between good habits and excellent academic achievement:

1. Attend classes regularly.
2. Take notes while teaching.
3. Concentrate on studying.
4. Study with a view to gaining meaning, not storing facts.
5. Prepare a schedule.
6. Follow the schedule.
7. Have appropriate rest periods.
8. Facing problems considering the home environment and planning.
9. Facing the challenges posed by the school environment.
10. Keep a daily update of the work done.

The statistical results in [39] confirm that the study habits index is indeed an increasing function of ability, as illustrated in Figure 2a. By applying methodologies such as those indicated by the authors in [44,45], one can adequately prepare a questionnaire including questions related to the preceding list.

On the other hand, a lack of academic and social skills leads to the student being unable to process the information transmitted by the instructor [46]. Then, we can infer that the understanding of topics is related to the student's ability, as Figure 2b illustrates.

By means of Figure 2a, we state that the study habits index behaves sigmoidally depending on the examinee's ability with the following correspondence rule

$$f(\theta) = B_h + \frac{1 - B_h}{1 + b_h e^{-a_h \theta}}, 0 < a_h. \quad (8)$$

Meanwhile, Figure 2b states that the rate of topic comprehension by students also has a sigmoidal behavior as follows

$$g(\theta) = B_c + \frac{1 - B_c}{1 + b_c e^{-a_c \theta}}, 0 < a_c. \tag{9}$$

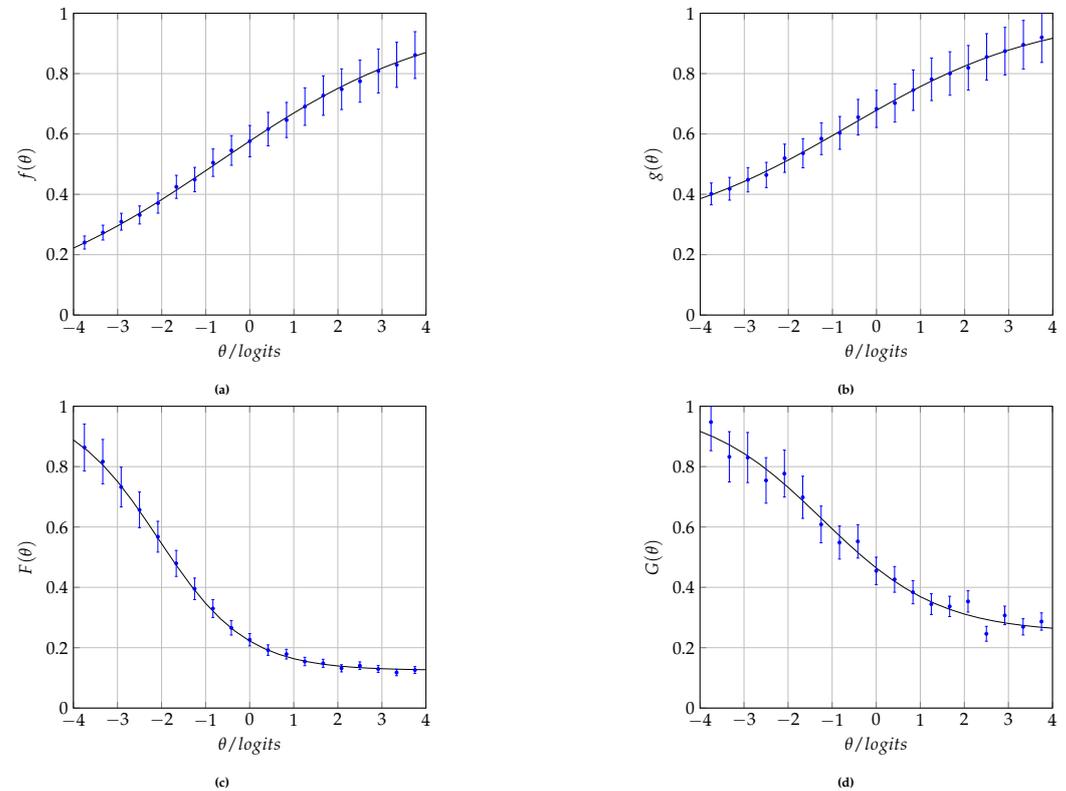


Figure 2. Assumed behavior for some of the indices that can be included as constraints for the determination of the prior distribution by means of entropy. (a) Study habit index f as a function of ability θ . (b) Subject comprehension index g as a function of ability θ . (c) Course dropout rate F as a function of ability θ . (d) Course failure rate G as a function of ability θ .

In order to be rational, we consider the good study habits rate in conjunction with the students’ failure rate as a function of the ability θ as follows:

1. For a randomly selected group of students, determine their abilities $\theta_1, \theta_2, \dots, \theta_n$.
2. For each of the selected examinees, as indicated in the former point 1, investigate the total number of failed subjects throughout their academic history.
3. With the assigned ability, the quotient of the total number of failed subjects and the total number of subjects taken or studied (considering even repetitions or recursing) define the failure rate for a specific examinee.

The third step is reinforced by the results published in [47], where they claim that low levels of ability tend to cause dropout from a course, if not from the school itself. The failure rate has an identical behavior and, for all these reasons, Figure 2c and Figure 2d, respectively, postulate that the dropout ($F(\theta)$) and failure ($G(\theta)$) from a course decrease exponentially with the ability of the student. The following correspondence rules illustrate these functions:

$$F(\theta) = 1 - \frac{1 - B_d}{1 + b_d e^{-a_d \theta}}, 0 < a_d, \tag{10}$$

$$G(\theta) = 1 - \frac{1 - B_r}{1 + b_r e^{-a_r \theta}}, 0 < a_r. \tag{11}$$

In all cases, note that the ability θ is a random variable and that the functions $f, g, F,$ and G are, therefore, random variables. In summary, the graphs in Figure 2 are the results of ansatzes here proposed to illustrate the behaviors of the random variables $f, g, F,$ and $G.$

Taking into account the postulated index functions and Equations (6) and (7), the Lagrangian \mathcal{L} to optimize, is given by Equation (12).

$$\begin{aligned} \mathcal{L}(p(\cdot)) = & - \int_{-\infty}^{+\infty} p(\theta) \log p(\theta) d\theta + \lambda_0 \left(\int_{-\infty}^{+\infty} p(\theta) d\theta - 1 \right) + \lambda_1 \left(\int_{-\infty}^{+\infty} \theta p(\theta) d\theta - \hat{\theta}_p \right) \\ & + \lambda_2 \left(\int_{-\infty}^{+\infty} (\theta - \hat{\theta}_p)^2 p(\theta) d\theta - \hat{\sigma}_p^2 \right) + \lambda_3 \left(\int_{-\infty}^{+\infty} \left(B_h + \frac{1 - B_h}{1 + b_h e^{-a_h \theta}} \right) p(\theta) d\theta - C_h \right) \\ & + \lambda_4 \left(\int_{-\infty}^{+\infty} \left(B_c + \frac{1 - B_c}{1 + b_c e^{-a_c \theta}} \right) p(\theta) d\theta - C_c \right) + \lambda_5 \left(\int_{-\infty}^{+\infty} \left(1 - \frac{1 - B_d}{1 + b_d e^{-a_d \theta}} \right) p(\theta) d\theta - C_d \right) \\ & + \lambda_6 \left(\int_{-\infty}^{+\infty} \left(1 - \frac{1 - B_r}{1 + b_r e^{-a_r \theta}} \right) p(\theta) d\theta - C_r \right), \end{aligned} \tag{12}$$

where Table A1 in Appendix A provides the meaning of the symbols appearing in the equations.

Without loss of generality and for reasons of simplicity, one only considers the constraint referring to the course failure rate $G(\theta)$ so that the maximization of entropy solves the set of non-linear equations defined by Equations (13)–(16) through:

$$\int_{-\infty}^{+\infty} e^{-1 + \lambda_0 + \lambda_1 \theta + \lambda_2 (\theta - \hat{\theta}_p)^2 + \lambda_6 \left(1 - \frac{1 - B_r}{1 + b_r e^{-a_r \theta}} \right)} d\theta - 1 = 0, \tag{13}$$

$$\int_{-\infty}^{+\infty} \theta e^{-1 + \lambda_0 + \lambda_1 \theta + \lambda_2 (\theta - \hat{\theta}_p)^2 + \lambda_6 \left(1 - \frac{1 - B_r}{1 + b_r e^{-a_r \theta}} \right)} d\theta - \hat{\theta}_p = 0, \tag{14}$$

$$\int_{-\infty}^{+\infty} (\theta - \hat{\theta}_p)^2 e^{-1 + \lambda_0 + \lambda_1 \theta + \lambda_2 (\theta - \hat{\theta}_p)^2 + \lambda_6 \left(1 - \frac{1 - B_r}{1 + b_r e^{-a_r \theta}} \right)} d\theta - \hat{\sigma}_p^2 = 0, \tag{15}$$

$$\int_{-\infty}^{+\infty} \left(1 - \frac{1 - B_r}{1 + b_r e^{-a_r \theta}} \right) e^{-1 + \lambda_0 + \lambda_1 \theta + \lambda_2 (\theta - \hat{\theta}_p)^2 + \lambda_7 \left(1 - \frac{1 - B_r}{1 + b_r e^{-a_r \theta}} \right)} d\theta - C_r = 0. \tag{16}$$

Note that Equation (6), both for the discrete (with summatory symbol instead of integral) and continuous cases can be considered as a measure of the misinformation (un-informativeness) that the prior distribution $p(\theta)$ provides about how the latent trait θ distributes [38,48]. This result is also supported by a research paper [48] where they state that the entropy maximization due to the constrained non-uniform prior distribution being equivalent to minimizing the distance between this distribution and an unconstrained uniform a priori distribution with no other constraint than a normalization process.

5. Results

Algorithm 2 illustrates our general procedure to select the a priori distribution. Line 1 of the algorithm assigns an initial estimation of the latent trait average $\hat{\theta}_p$ and variance average $\hat{\sigma}_p^2$. Additionally, the performance index function structures are defined as closely related to the experimental procedures (academic) already mentioned. Finally, lines 5 to 8 find the conditions to properly select an a priori distribution satisfying the normalization, ability’s average value, variance’s average value, and average values of expected performance index functions.

There are assumptions or inconveniences that appear when one applies a MAP or EAP technique at the initialization of the CAT process, considering that one does not know something about the prior distribution when the CAT system provides the first item. However, fortunately, there are several ways to solve this problem [8,35,49]. In this work, one proposes an initial ability equal to the value $\hat{\theta}_p$ given to the constraint in Equation (14). Thus we can apply Algorithm 3 to simulate the CAT process.

Algorithm 2 Diagram of the prior distribution search process.

```

1: procedure SEARCHPRIOR
2:    $\theta_p^* \leftarrow \hat{\theta}_p$  ▷ assign average of given skill in restriction (14)
3:    $\sigma_p^{2*} \leftarrow \hat{\sigma}_p^2$  ▷ assigns expected variance in constraint (15)
4:    $\vec{p} \leftarrow$  defines index parameters used in constraints
5: top:
6:    $s \leftarrow$  solve system of nonlinear equations defined by constraints
7:   if unsatisfied constraints then
8:     goto top
9:   return  $\vec{l}$  ▷ returns Lagrange multipliers

```

Algorithm 3 Scheme of the CAT process using Bayesian estimation with prior distribution.

```

1: procedure ADAPTABLEEVALUATIONPROCESSWITHPRIOR
2:    $responses \leftarrow []$ 
3:    $\theta^* \leftarrow \hat{\theta}_p$  ▷ allocates average of given skill in constraint (14)
4:    $p(\theta) \leftarrow$  determine prior distribution maximizing entropy with constraints
5: top:
6:    $i \leftarrow \max_{Pool} \{I(\theta^*)\}$ 
7:    $r \leftarrow$  reply to item  $i, 1 \leq i$ 
8:    $responses \leftarrow concat(responses, r)$  ▷ update response history
9:    $L(\vec{p}|\theta) \leftarrow$  determines the likelihood function as a product of ICCs
10:  if  $responses$  are all correct or incorrect then ▷ uses Bayesian inference
11:     $p(\theta|\vec{p}) \leftarrow L(\vec{p}|\theta)p(\theta)$ 
12:     $p(\theta|\vec{p}) \leftarrow kp(\theta|\vec{p})$  ▷ normalizes posterior distribution
13:     $\theta^* \leftarrow \int_{-\infty}^{+\infty} \theta p(\theta|\vec{p}) d\theta$  ▷ compute average skill with new distribution
14:     $p(\theta) \leftarrow p(\theta|\vec{p})$ 
15:  else
16:     $\theta^* \leftarrow$  use Maximum Likelihood estimation as usual
17:  goto top.
18:  return  $\theta^*$ 

```

After running a sequence of simulations under the directions of Algorithm 3, one obtains as examples the corresponding CAT processes that Figures 3–5 show. Table 1 gives some numerical results, whereas the third experiment shows the complete running. The following list synthesizes the obtained results of the corresponding simulation process.

1. After several iterations, the CAT system always tends to the maximum Fisher's information index, regardless of the intermediate value of the estimated ability θ . Thus, the final selected item has a difficulty μ (see Figures 3–5).
2. When the study habits index function discriminates well and plays the role of one constraint in entropy maximization, one can expect a bimodal a priori distribution as acceptable (see Figure 3)
3. A possible behavior in the initialization of the CAT process when the discriminating power of the study habits index function is not high or low can be found in Figure 4. Note that the a priori distribution shows some non-null skewness.
4. A failure rate with a lower discrimination index provides an initial prior distribution with almost null skewness. So, in practice, when one takes a normal or Gaussian prior distribution $\mathcal{N}(\theta; \mu, \sigma)$ with a high variance σ^2 [50] or a uniform distribution $\mathcal{U}(a, b), a \ll b$, one also assumes that the examinees' failure records are the same.

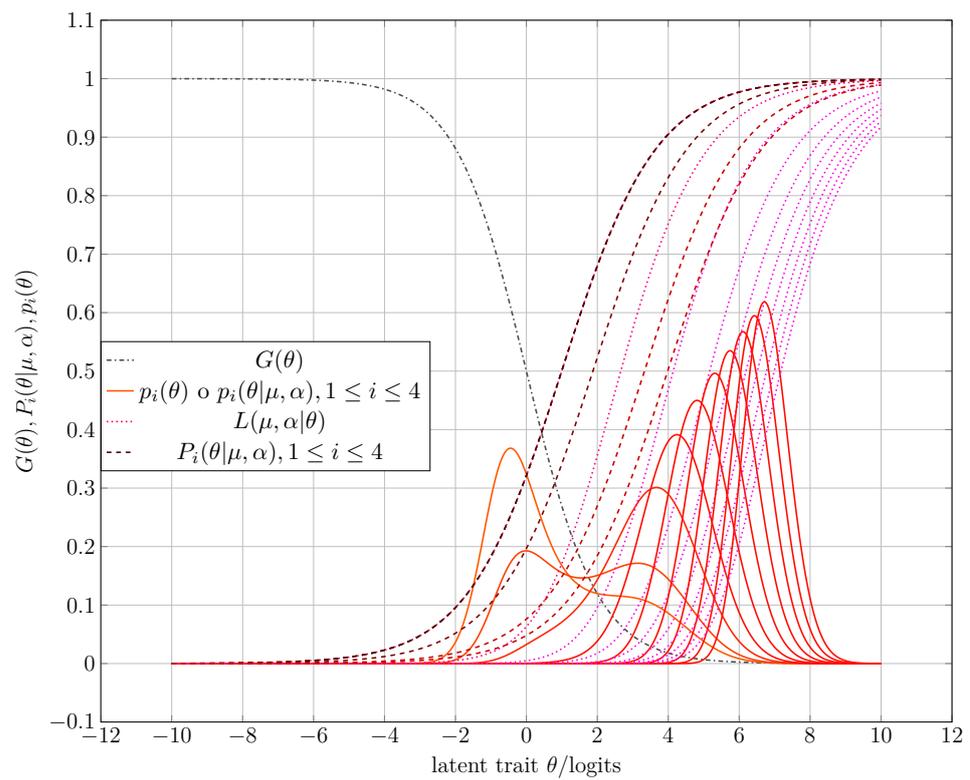


Figure 3. Starting of the CAT for the first experiment. Table 1 shows the expected and computed parameters from entropy maximization and the simulation process.

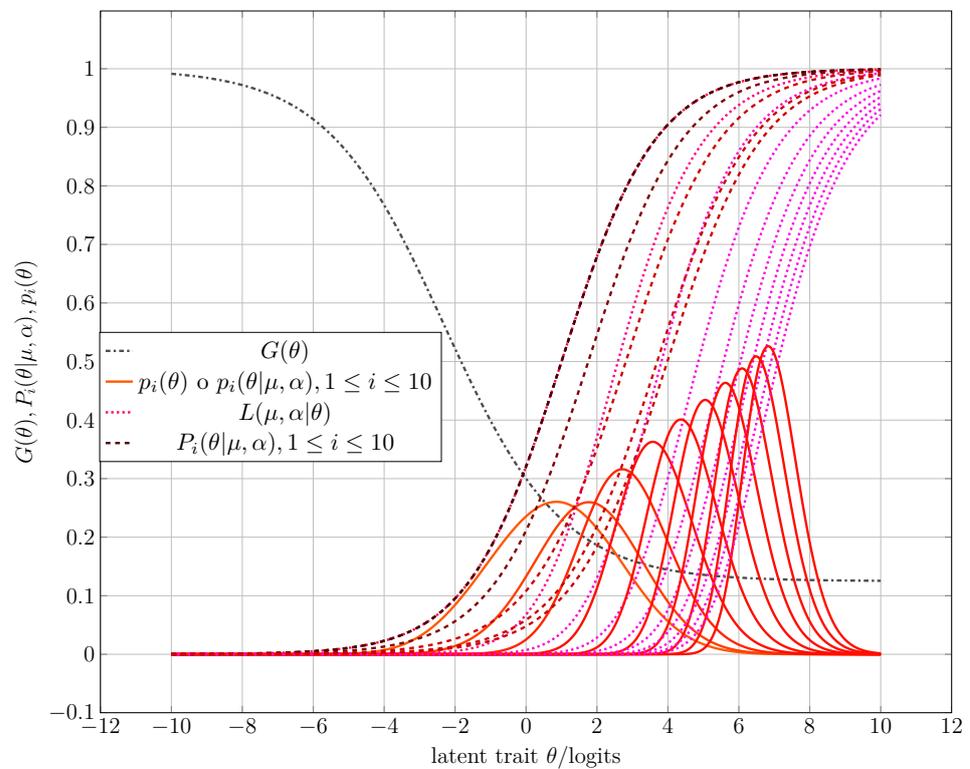


Figure 4. Starting of the CAT for the second experiment. Table 1 shows the expected and computed parameters from entropy maximization and the simulation process.

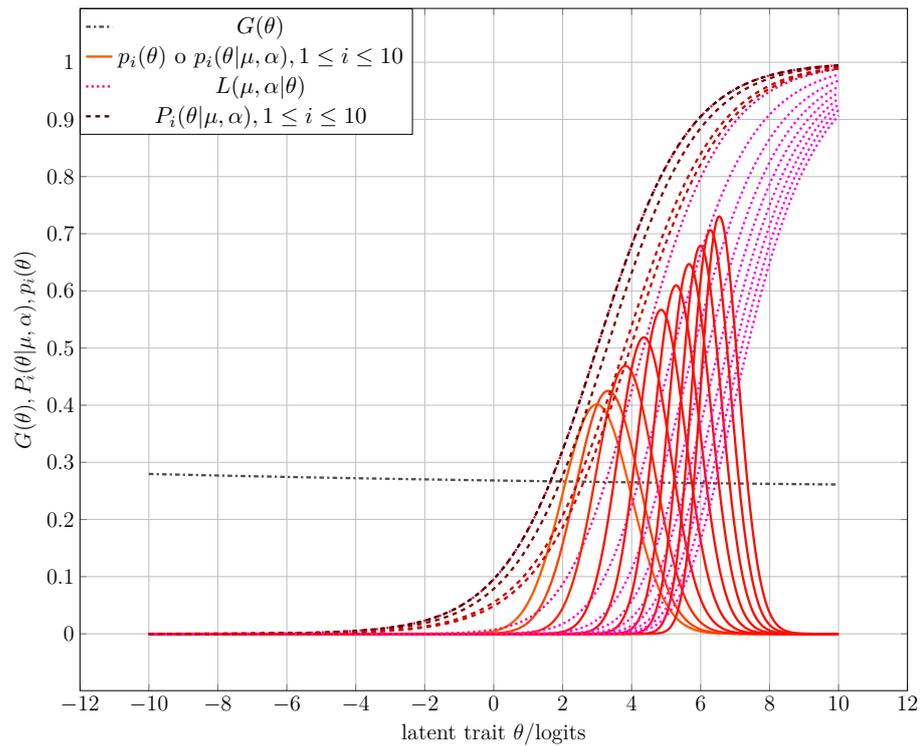


Figure 5. Starting of the CAT for the third experiment. Table 1 shows the expected and computed parameters from entropy maximization and the simulation process.

Table 1. Results of numerical experimentation. The parameters values of the Start Parameters column assume that they come from estimations and/or experiments. The parameters values of the prior properties column assume that they come from the computation of the corresponding integral expression containing the computed prior distribution (first terms in left-hand sides of Equations (13)–(16)). The numerical experimentation of a running for a CAT in Experiment 3 assumes the existence of a pool of 1000 calibrated items.

Experiment	Start Parameters						prior properties				skewness	Lagrange’s multipliers			
1	θ_p^*	σ_p^{*2}	a_r	b_r	B_r	C_r^*	normality	θ	σ^2	C_r		λ_1	λ_2	λ_3	λ_4
	1	4	1	1	0	0.6	1.18	0.98	3.99	0.47	3.66	−6.60	2.12	−0.40	13.74
Experiment	Start Parameters						Properties prior				skewness	Lagrange’s multipliers			
2	θ_p^*	σ_p^{*2}	a_r	b_r	B_r	C_r^*	normality	θ	σ^2	C_r		λ_1	λ_2	λ_3	λ_4
	1	4	0.6	0.25	0.125	0.6	1.20	0.96	3.99	0.34	−2.00	−1.34	0.16	−0.19	3.62
Experiment	Start Parameters						Properties prior				skewness	Lagrange’s multipliers			
3	θ_p^*	σ_p^{*2}	a_r	b_r	B_r	C_r^*	normality	θ	σ^2	C_r		λ_1	λ_2	λ_3	λ_4
	3	1	0.05	0.025	0.25	0.3	1.00	3.00	1.00	0.27	−0.04	−3.49	0.00	−0.5	13.46
1000 difficulties in items Pool															
	iteration		θ_p^*		μ_p^*		normality		map		eap				
	1		3		2.9978		1.00411027		2.99		3.00159777				
	2		3.3222		3.3207		1.00		3.32		3.32218745				
	3		3.8357		3.8430		1.00		3.82		3.83572950				
	4		4.3893		3.9851		1.00		4.36		4.38926317				
	5		4.8986		3.9851		1.00		4.85		4.89863112				
	6		5.3412		3.9851		1.00		5.29		5.34123495				
	7		5.7224		3.9851		1.00		5.67		5.72237081				

There are several fine details to work out when one uses a priori distributions [51]; however, in this paper, we provide a unified approach to derive prior distributions with a

less subjective selection of the distribution when the initialization of the CAT process uses Bayesian estimation [52,53].

There is a large number of research papers published about the advantages and drawbacks on the use of a priori distributions topic, but the techniques used there are based on heuristics to build the Bayesian inference procedure within the initialization of the CAT process in some special cases [51,54].

In order to compare likely differences between the results of heuristic techniques and our methodology, a useful tool to be used is the Kullback–Leibler (KL) divergence index. This index measures the divergence of the expected amount of extra information required to obtain population samples that follow the prior distribution $p(\theta)$ when using population samples that follow a distribution $q(\theta)$ [55].

The KL divergence measure is defined by Equation (17).

$$D_{KL}(p(\theta)||q(\theta)) = \int_{-\infty}^{+\infty} p(\theta) \ln \frac{p(\theta)}{q(\theta)} d\theta. \tag{17}$$

In this sense, the information is more ordered when one applies the prior distribution obtained with our method than with the popular unconstrained heuristic distributions. One should expect this result since the introduction of constraints orders the information under analysis. In this manner, $p(\theta)$ represents a “realistic” data distribution or a precisely calculated theoretical distribution and the typical distribution $q(\theta)$ represents a description or approximation of $p(\theta)$ (see [55]).

Through Table 1, a correspondence rule for the a priori distribution $p(\theta)$ can be defined such as Table 2 illustrates. Therefore, the two measures given by Equations (18) and (19) considering the normal and uniform distributions, respectively, can be calculated.

$$D_{KL}(p(\theta)||\mathcal{N}(\theta; \mu, \sigma)) = \int_{-\infty}^{+\infty} p(\theta) \ln \frac{p(\theta)}{\mathcal{N}(\theta; \mu, \sigma)} d\theta, \tag{18}$$

$$D_{KL}(p(\theta)||\mathcal{U}(\theta; a, b)) = \int_{-\infty}^{+\infty} p(\theta) \ln \frac{p(\theta)}{\mathcal{U}(\theta; a, b)} d\theta. \tag{19}$$

Table 2. A priori distributions distances with respect to heuristics ones for every experiment in Table 1.

Experiment	$p(\theta)$	$q(\theta)$	KL Distance
1	$e^{-1+\lambda_1+\lambda_2\theta+\lambda_3(\theta-\theta_p^*)^2+\lambda_4\left(1-\frac{1-Br}{1+br e^{-ar\theta}}\right)}$	$\mathcal{N}(\theta; 1, 2)$	3.5402
		$\mathcal{U}(\theta; 1 - 2.225, 1 + 2.225)$	0.73339
2	$e^{-1+\lambda_1+\lambda_2\theta+\lambda_3(\theta-\theta_p^*)^2+\lambda_4\left(1-\frac{1-Br}{1+br e^{-ar\theta}}\right)}$	$\mathcal{N}(\theta; 1, 2)$	2.4119
		$\mathcal{U}(\theta; 1 - 2.385, 1 + 2.385)$	0.077474
3	$e^{-1+\lambda_1+\lambda_2\theta+\lambda_3(\theta-\theta_p^*)^2+\lambda_4\left(1-\frac{1-Br}{1+br e^{-ar\theta}}\right)}$	$\mathcal{N}(\theta; 3, 1)$	0.069902
		$\mathcal{U}(\theta; 3 - 1.7375, 3 + 1.7375)$	0.084054

From Table 2, note that the distance given by the KL divergence in the first experiment when comparing the prior distribution with the Gaussian distribution $\mathcal{N}(\theta; \mu, \sigma)$ suggests that to analyze the population with this last distribution, one should expect an amount of 3.5402 extra information to include the data population related to the first distribution. Figure 6a–c compare the three distributions for every experiment in Table 2, and show their respective KL measures.

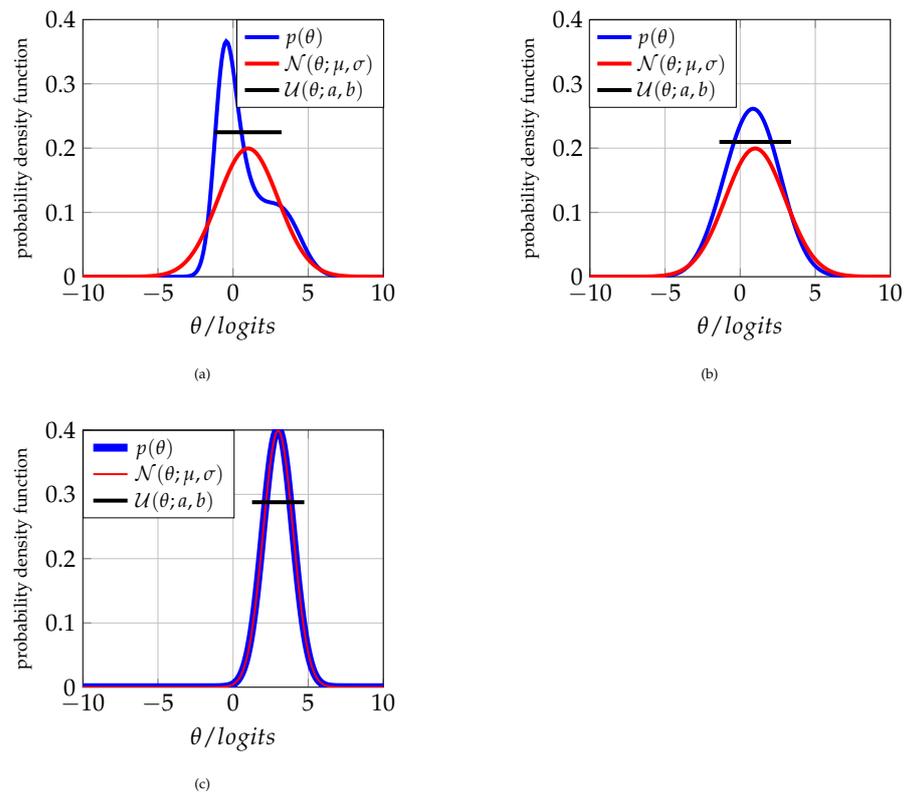


Figure 6. KL distance between each of the initial prior distributions $p(\theta)$ from each experiment and the classic distributions $\mathcal{N}(\theta; \mu, \sigma)$ and $\mathcal{U}(\theta; a, b)$. (a) First experiment: $p(\theta)$ vs. $\mathcal{N}(\theta; 1, 2)$ and $\mathcal{U}(\theta; a, b)$. (b) Second experiment: $p(\theta)$ vs. $\mathcal{N}(\theta; 1, 2)$ and $\mathcal{U}(\theta; a, b)$. (c) Third experiment: $p(\theta)$ vs. $\mathcal{N}(\theta; 3, 1)$ and $\mathcal{U}(\theta; a, b)$.

In the first experiment, when one compares the distribution $\mathcal{U}(\theta; 1 - 2.225, 1 + 2.225)$, the a priori distribution $p(\theta)$ also has a KL divergence equal to 0.73339. On the other hand, for the second experiment, KL divergences become equal to 2.4119 and 0.077474 when one, respectively, approximates through $\mathcal{N}(\theta; 1, 2)$ and $\mathcal{U}(\theta, 1 - 2.385, 1 + 2.385)$. Finally, for the third experiment, the KL divergences equal to 0.069902 and 0.084054 when one, respectively, approximates through $\mathcal{N}(\theta; 3, 1)$ and $\mathcal{U}(\theta; 3 - 1.7375, 3 + 1.7375)$.

Intervals (a, b) for every uniform distribution are calculated by looking for the lower distance between the corresponding $p(\theta)$ and $\mathcal{U}(\theta; a, b)$ distributions. Note that the third experiment results agree with the heuristic suggestion of using the normal, or uniform distributions, as good approximations to the prior distribution. So, the alternative is acceptable when the course failure index function does not discriminate well.

6. Conclusions

In this paper, we demonstrate that through the theory of entropy maximization, a given set of constraints, and under numerical experimentation, the computation of an a priori distribution to initialize a CAT process by using Bayesian inference can be carried out. Furthermore, the examinee’s performance index functions define the constraints, and they complement the usual distribution constraints (normality, first and second moment, etc.).

We also show that through the entropy theory, the selection of appropriate constraints summarizes experimental data through the specification of index functions related to study habits, comprehension levels, course dropout, and lecture failure.

A given set of constraints can produce a set of acceptable or unacceptable a priori distributions, so one needs to look for a stop criterion in searching for the optimal set of parameters that defines the distributions through entropy maximization. To define the stop criterion, we verify how the estimated set of distribution parameters and those that define the constraints are close enough to the expected values used in the constraints

definition. Thus, the most appropriate distribution is chosen and, under the assumption of responding correctly to the first items in the testing process, we can verify its latent trait prediction capability.

Index functions playing the role of constraints with acceptable discrimination properties produce a priori distributions with bimodality, as one can expect, so that the obtained distribution estimates reasonable latent trait values along the simulation of the CAT process.

In summary, entropy maximization can be used inside the frame of a CAT to derive more generalized a priori distributions through constraint specifications related to index functions. This method can provide a unified approach to derive a priori distributions for initializing the CAT process through a Bayesian inference procedure.

Author Contributions: Conceptualization, J.S.-C.; Methodology, J.S.-C. and V.L.-M.; Software, J.S.-C. and L.R.M.-M.; Validation, J.S.-C.; Formal analysis, J.S.-C. and V.L.-M.; Investigation, V.L.-M., L.R.M.-M., A.A.-R. and J.C.R.-F.; Resources, A.A.-R.; Data curation, J.C.R.-F.; Writing—original draft, V.L.-M.; Writing—review & editing, V.L.-M.; Visualization, A.A.-R. and J.C.R.-F.; Supervision, V.L.-M. All the authors have equally contributed and worked in this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like, through these lines, to really thank the anonymous reviewers and the journal’s staff for their valuable support to improve the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

1PL	One-Parameter Logistic
2PL	Two-Parameter Logistic
CAT	Computer Adaptive Testing
CIP	Calibrated Item Pool
EAP	Expectation a posteriori
IRT	Item Response Theory
ICC	Item Characteristic Curve
KL	Kullback–Leibler
MAP	<i>maximum a posteriori</i>
w.r.to	with respect to

Appendix A. Table of Symbols

Table A1. Summary of all the symbols used in the paper, their location and meaning.

Symbol	Meaning	Pages Where the Symbol Appears
e	Base of the natural logarithm	2
θ	Examinee’s ability or latent trait to be estimated in the CAT process	2–6, 10, 12–16
θ_i	Examinee’s ability at i -th iteration	6
μ	Item’s difficulty playing the role of a parameter in the definition of the ICC that the CDF P defines	2, 5, 6, 15, 16
$\mu_i, i = 1, 2, \dots$	Difficulty of item i	5
α	Item’s discriminant playing the role of a parameter in the definition of the ICC that the PDF P defines	5, 6, 13, 14

Table A1. Cont.

Symbol	Meaning	Pages Where the Symbol Appears
P	Conditional cumulative distribution function for computing the probability that the examinee with ability θ gives a correct answer to an item, given that the item's difficulty is μ and (possibly) the discriminant is α	2, 5, 6
Q	It is equal to $1 - P$; in other words, it represents the probability that the examinee with ability θ gives an incorrect answer to an item, given that the item's difficulty is μ and (possibly) the discriminant is α	6
$\Delta\mu$	Increment of item's difficulty	6
$\Delta\alpha$	Increment of item's discriminant	5
$n = 1, 2, \dots$	Number of items at iteration n , or n -th iteration, in the CAT process	6
I	Fisher's information, as a function of the examinee's ability θ , is useful to compute the amount of information that a given item provides about the θ estimate by knowing the current value of θ	6
\vec{p}	Vector of parameters (item's difficulty, item's discriminant, etc.) that defines the model P	6, 12
p	If p depends on θ , then it represents a prior distribution; otherwise, if p depends on θ given the vector of parameters \vec{p} , then it represents a posterior distribution	7, 16
L	Likelihood function depends on the vector of parameters \vec{p} given the examinee's ability θ , $L(\vec{p}_i, i = 1, 2, \dots, n \theta) \propto \prod_{i=1}^n P_i^k(\theta \vec{p}_i) Q_i^{n-k}(\theta \vec{p}_i)$, where k represents the number of correct answers that an examinee gives to n items	7
f, g, F, G	Performance indexes are functions of a random variable X	7
S	Entropy	7
\mathcal{L}	Lagrangian	8, 11
λ	Lagrange multiplier	8, 11
$B_i, b_i, a_i,$ where $i = h, c, d, r$	Parameters defining performance index function: study habit, subject comprehension, course dropout rate, and course failure rate, respectively	10
$\hat{\theta}_p$	Initial estimation of the latent trait θ , just before starting the CAT process, the subindex p comes from the word prior	11, 12
$\hat{\sigma}_p^2$	Initial estimation of the variance σ^2 of the distribution of the latent trait θ , the subindex p comes from the word prior	11, 12
s	It contains the solution to entropy maximization under the given constraints	12
\vec{l}	Vector of Lagrange multipliers after satisfying entropy maximization under given constraints	12

Table A1. Cont.

Symbol	Meaning	Pages Where the Symbol Appears
θ^*	Initial estimation of the latent trait θ , just before starting the CAT process, or expectation a posteriori of the latent trait along the CAT process, or the extreme point of the ability values θ , along the CAT process, where the a posteriori distribution has a <i>maximum</i>	12, 13
\mathcal{U}	Uniform distribution	13, 15, 16
\mathcal{N}	Normal distribution	13, 15, 16
a, b with $a < b$	lower and upper bounds, respectively, that define the uniform distribution	15, 16
σ^2, σ	Variance and standard deviation estimation, respectively, for a priori distribution	13, 15
C_r	Estimation of the course failure rate average by means of the estimated a priori distribution	11, 13
C_r^*	Course failure rate estimation coming from the experimental results for obtaining the course failure rate index function	11, 13

References

- Haifeng, L.; Ning, Z.; Zhixin, C. A Simple but Effective Maximal Frequent Itemset Mining Algorithm over Streams. *J. Softw.* **2012**, *7*, 25–32.
- Li, M.; Han, M.; Chen, Z.; Wu, H.; Zhang, X. FCHM–Stream: Fast Closed High Utility Itemsets Mining over Data Streams. Research Article, Research Square. 2022, 19p. Available online: https://assets.researchsquare.com/files/rs-1736816/v1_covered.pdf?c=1655222833 (accessed on 14 December 2022).
- Liu, J.; Ye, Z.; Yang, X.; Wang, X.; Shen, L.; Jiang, X. Efficient strategies for incremental mining of frequent closed itemsets over data streams. *Expert Syst. Appl.* **2022**, *191*, 116220. [CrossRef]
- Caruccio, L.; Cirillo, S.; Deufemia, V.; Polese, G. Efficient Discovery of Functional Dependencies from Incremental Databases. In Proceedings of the 23rd International Conference on Information Integration and Web Intelligence, IIWAS2021, Linz, Austria, 29 November–1 December 2021; pp. 400–409.
- Hu, K.; Qiu, L.; Zhang, S.; Wang, Z.; Fang, N. An incremental rare association rule mining approach with a life cycle tree structure considering time-sensitive data. *Appl. Intell.* **2022**. [CrossRef]
- Revelle, W. Chapter 8 The “New Psychometrics”– Item Response Theory. In *An Introduction to Psychometric Theory with Applications in R*; 2013; pp. 241–264. Available online: <http://personality-project.org/courses/405.syllabus.html> (accessed on 14 December 2022).
- DeMars, C. *Item Response Theory*; Oxford University Press: Cary, NC, USA, 2010.
- Olea, J.; Ponsoda, V. *Tests Adaptativos Informatizados*; Universidad Nacional de Educación a Distancia (UNED) Ediciones: Madrid, Spain, 2003.
- Revuelta, J.; Ponsoda, V. Una Solución a la estimación inicial en los Tests Adaptativos Informatizados. *Rev. Electrónica De Metodol. Apl.* **1997**, *2*, 1–6.
- Frans, N.; Braeken, J.; Veldkamp, B.P.; Paap, M.C.S. Empirical Priors in Polytomous Computerized Adaptive Tests: Risks and Rewards in Clinical Settings. *Appl. Psychol. Meas.* **2022**, *47*, 48–63. [CrossRef]
- O’Hagan, A.; Luce, B.R. *A Primer on BAYESIAN STATISTICS in Health Economics and Outcomes Research*; Bayesian Initiative in Health Economics & Outcomes Research, Center for Bayesian Statistics in Health Economics, MEDTAP International: Bethesda, MD, USA, 2003.
- Veldkamp, B.P.; Matteucci, M. Bayesian Computerized Adaptive Testing. *Ensaio Avaliação e Políticas Públicas em Educação* **2013**, *21*, 57–82. [CrossRef]
- Liu, X.; Lu, D. A MAP method with nonparametric priors for estimating P–S–N curves. In Proceedings of the Fifth International Symposium on Life–Cycle of Engineering Systems: Emphasis on Sustainable Civil Infrastructure Conference, Delft, The Netherlands, 16–19 October 2016; pp. 2120–20124.
- Swaminathan, H.; Gifford, J.A. Bayesian estimation in the three-parameter logistic model. *Psychometrika* **1986**, *51*, 589–601. [CrossRef]
- Wang, T.; Vispoel, W.P. Properties of ability estimation methods in computerized adaptive testing. *J. Educ. Meas.* **1998**, *35*, 109–135. [CrossRef]

16. Lord, F.M. Maximum likelihood and Bayesian parameter estimation in item response theory. *J. Educ. Meas.* **1986**, *23*, 157–162. [[CrossRef](#)]
17. Samejima, F. Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychom. Monogr.* **1969**, *17*, 100. Available online: <http://www.psychometrika.org/journal/online/MN17.pdf> (accessed on 14 December 2022). [[CrossRef](#)]
18. Mitrushina, M.; Boone, K.B.; Razani, J.; D'Elia, L.F. Statistical and Psychometric Issues. In *Handbook of Normative Data for Neuropsychological Assessment*, 2nd ed.; Mitrushina, M., Boone, K.B., Razani, J., D'Elia, L.F., Eds.; Oxford University Press: New York, NY, USA, 2005; pp. 33–56.
19. Ho, A.D.; Yu, C.C. Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educ. Psychol. Meas.* **2015**, *75*, 365–388. [[CrossRef](#)] [[PubMed](#)]
20. Mitrushina, M. *Handbook of Normative Data for Neuropsychological Assessment*; Oxford University Press: New York, NY, USA, 2005; pp. 38–39.
21. Stephens, M. Dealing with multimodal posteriors and non-identifiability in mixture models. *J. R. Stat. Soc. Ser. B* **1999**, *62*, 795–809. Available online: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c0d3e6bf574ec653f5b81d39149413c7a3aa949a> (accessed on 14 December 2022). [[CrossRef](#)]
22. Botje, M. *Introduction to Bayesian Inference. Lecture Notes at NIKHEF National Instituut Voor Subatomaire Fysica*; Publisher National Instituut voor Subatomaire Fysica: Amsterdam, The Netherlands, 2006.
23. LaValle, S.M. *Planning Algorithms*; Cambridge University Press: Cambridge, UK, 2006.
24. Bromiley, P.A.; Thacker, N.A.; Bouhova-Thacker, E. *Shannon Entropy, Renyi Entropy, and Information*; Technical Report No. 2004-004; Imaging Science and Biomedical Engineering, School of Cancer and Imaging Science: London, UK, 2010.
25. Bretthorst, G.L. An Introduction to Parameter Estimation Using Bayesian Probability Theory. In *Maximum Entropy and Bayesian Methods*; Kluwer Academic, P.F.F., Ed.; Springer: Dordrecht, The Netherlands, 1990; pp. 53–79.
26. Jaynes, E.T. Prior Probabilities. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 227–241. [[CrossRef](#)]
27. Jeon, S.S.; Su, S.Y.W. Deriving Prior Distributions for Bayesian Models Used to Achieve Adaptive E-Learning. *Knowl. Manag. E-Learn. Int. J. (KM EL)* **2011**, *3*, 251–270.
28. Albert, I.; Donnet, S.; Guihenneuc-Jouyaux, C.; Low-Choy, S.; Mengersen, K.; Rousseau, J. Combining Expert Opinions in Prior Elicitation. *Bayesian Anal.* **2012**, *7*, 503–532. [[CrossRef](#)]
29. Dayanik, A.; Lewis, D.D.; Madigan, D.; Menkov, V.; Genkin, A. Constructing informative prior distributions from domain knowledge in text classification. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, USA, 6–11 August 2006; pp. 493–500.
30. Keller, L.A. *Ability Estimation Procedures in Computerized Adaptive Testing*; Technical Report; American Institute of Certified Public Accountants: Durham, NC, USA, 2000.
31. Tchourbanov, A. *Prior Distributions*; Technical Report; Department of Biology New Mexico State University Road Runner Gnomics Laboratories. 2002. Available online: [https://datajobs.com/data-science-repo/Conjugate-Priors-\[Alexandre-Tchourbanov\].pdf](https://datajobs.com/data-science-repo/Conjugate-Priors-[Alexandre-Tchourbanov].pdf) (accessed on 14 December 2022).
32. Gelman, A.; Simpson, D.; Betancourt, M. The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy* **2017**, *19*, 555. [[CrossRef](#)]
33. Matteucci M.; Veldkamp, B.P. Bayesian Estimation of Item Response Theory Models with Power Priors. In Proceedings of the Statistical Conference, Advances in Latent Variables, Methods, Models and Applications, SIS 2013, Brescia University, Owensboro, KY, USA, 19–21 July 2013. (Special Session)
34. Muñiz, J. Introducción a la Teoría de Respuesta a los Ítems. Ediciones Pirámide, Colección Psicología, Sección Psicometría, Madrid. 1997. Available online: <https://www.semanticscholar.org/paper/Introducci%C3%B3n-a-la-teor%C3%ADa-de-respuesta-a-los-%C3%ADtems-Fern%C3%A1ndez/4bd320747e2df2f34ba71e61199ff49e93df007e> (accessed on 14 December 2022).
35. van der Linden, W.J.; Pashley, P.J. Chapter 1 Item Selection and Ability Estimation in Adaptive Testing. In *Computerized Adaptive Testing: Theory and Practice*; van der Linden, W.J., Glas, C.A.W., Eds.; Springer: Dordrecht, The Netherlands, 2000; pp. 1–25.
36. Yao, L. Item Selection Methods for Computer Adaptive Testing With Passages. *Front. Psychol.* **2019**, *10*, 240. [[CrossRef](#)]
37. Olea, J.; Ponsoda, V. Capítulo 4 Algoritmos Adaptativos. In *Tests Adaptativos Informatizados*; Universidad Nacional de Educación a Distancia (UNED) Ediciones: Madrid, Spain, 2003; pp. 47–66.
38. Consonni, G.; Fouskakis, D.; Liseo, B.; Ntzoufras, I. Prior Distributions for Objective Bayesian Analysis. *Bayesian Anal.* **2018**, *13*, 627–679. [[CrossRef](#)]
39. Siah, E.A.; Maiyo, J.K. Study of the relationship between study habits and academic achievement of students: A case of Spicer Higher Secondary School, India. *Int. J. Educ. Adm. Policy Stud.* **2015**, *7*, 134–141.
40. Ebele, U.F.; Olofu, P.A. Study habit and its impact on secondary school students' academic performance in biology in the Federal Capital Territory, Abuja. *Educ. Res. Rev.* **2017**, *12*, 583–588. [[CrossRef](#)]
41. Andrich, D. A Structure of Index and Causal Variables. *Trans. Rasch Meas. SIG Am. Educ. Res. Assoc.* **2014**, *28*, 1475–1477.
42. Andrich, D.; Marais, I. Chapter 4 Reliability and Validity in Classical Test Theory. In *A Course in Rasch Measurement Theory, Springer Texts in Education, Measuring in the Educational, Social and Health Sciences*; Berlin/Heidelberg, Germany, 2019; pp. 41–53. [[CrossRef](#)]
43. Stenner, A.J.; Stone, M.H.; Burdick, D.S. Indexing vs. Measuring. *Rasch Meas. Trans.* **2009**, *22*, 1176–1177. Available online: <https://www.rasch.org/rmt/rmt224b.htm> (accessed on 14 December 2022).

44. Ramakrishnan, S.; Robbins, T.W.; Zmigrod, L. Research Article—The Habitual Tendencies Questionnaire: A tool for psychometric individual differences research. *Personal. Ment. Health* **2022**, *16*, 30–46. [[CrossRef](#)] [[PubMed](#)]
45. Abed, B.K. Study Habits Used by Students at the University of Technology. *J. Educ. Coll.* **2016**, *1*, 537–558. Available online: <https://www.iasj.net/iasj/article/113575> (accessed on 14 December 2022).
46. Eleby, C. The Impact of a Student's Lack of Social Skills on their Academic Skills in High School. Master's Thesis, Marygrove College, Detroit, MI, USA, April 2009.
47. Cruz-Sosa, E.M.; Gática-Barrientos, L.; García-Castro, P.E.; Hernández-García, J. Academic Performance, School Desertion And Emotional Paradigm In University Students. *Contemp. Issues Educ. Res.* **2010**, *3*, 25–36. [[CrossRef](#)]
48. Kapur, J.N. Chapter 1 Maximum–Entropy Probability Distributions: Principles, Formalism and Techniques. In *Maximum–entropy Models in Science and Engineering*; Jagat Narain Kapur (Revised Edition); Wiley: Hoboken, NJ, USA, 1993; pp. 1–29.
49. Meijer, R.R.; Nering, M.L. Computerized Adaptive Testing: Overview and Introduction. *Appl. Psychol. Meas.* **1999**, *23*, 187–194. [[CrossRef](#)]
50. Raïche, G.; Blais, J.G.; Magis, D. Adaptive estimators of trait level in adaptive testing: Some proposals. In Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing, 8 June 2007; Weiss, D.J., Ed. Available online: https://www.researchgate.net/publication/228435165_Adaptive_estimators_of_trait_level_in_adaptive_testing_Some_proposals (accessed on 14 December 2022).
51. Cengiz, M.A.; Öztük, Z. A Bayesian Approach for Item Response Theory in Assessing the Progress Test in Medical Students. *Int. J. Res. Med. Health Sci.* **2013**, *3*, 15–19.
52. Van der Linden, W.J. Empirical Initialization of the Trait Estimator in Adaptive Testing. *Appl. Psychol. Meas.* **1999**, *23*, 21–29. [[CrossRef](#)]
53. Chen, L.; Singh, V.P. Entropy–based derivation of generalized distributions for hydrometeorological frequency analysis. *J. Hydrol.* **2018**, *557*, 699–712. [[CrossRef](#)]
54. Gelman, A. Objections to Bayesian statistics. *Bayesian Anal.* **2008**, *3*, 445–450. [[CrossRef](#)]
55. Han, J. Chapter 2 Know Your Data (Additional Material) Kullback–Leibler Divergence. Lecture Notes (3rd ed.) CS412 Fall 2008 Introduction to Data Warehousing and Data Mining at the Department of Computer Science, University of Illinois, August 2017. Available online: <http://hanj.cs.illinois.edu/cs412/bk3/KL-divergence.pdf> (accessed on 14 December 2022).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.