To cite this article:

Tuerlinckx, F., & Robledo Araya, H. A. (in press). Geodesic ability in unidimensional IRT models: A geometric perspective on parametrization invariance. *Psychometrika*.

Geodesic ability in unidimensional IRT models: A geometric perspective on parametrization invariance

Francis Tuerlinckx[1] and Hernán Andrés Robledo Araya[2]

[1]University of Leuven, Leuven, Belgium

[2]Faculty of Mathematics, Pontificia Universidad Católica de Chile, Chile

Abstract

Psychometric models use latent variables to quantify psychological traits, but statistical and substantive inferences based on these variables may depend on how the latent scale is parameterized. Consequently, the same fitted model can yield qualitatively different conclusions under different parameterizations. Building on Ramsay's (1996) geometric perspective, we reinterpret this problem using differential geometry. While Ramsay defined ability via an arc length in Euclidean space, psychometric models can be seen instead as a curved statistical manifold with the Fisher–Rao metric. This allows for a parameterization-invariant definition of ability, which we term geodesic ability. We illustrate this concept for unidimensional IRT models, including the Rasch and 3PL models, and compare it to Ramsay's arc-length approach.

Geodesic ability in unidimensional IRT models: A geometric perspective on parametrization invariance

1 Introduction

Suppose two doctors are studying children's height. One doctor records height in meters, whereas the other uses inches. Although their numerical measurements differ, they will arrive at the same substantive conclusions, because height is directly observable (i.e., an extensive attribute; see Briggs, 2021) and can be compared to a unit. When different units are used, the measured quantities are related through a linear transformation.

Now consider two educational psychologists studying children's mathematical ability. Both psychologists use the same set of mathematics items and the same item response theory (IRT) model with known item parameters. The situation is profoundly different compared to the measurement of height: Even when the fitted model yields the same predicted response probabilities, different parametrizations of the latent scale can lead to different conclusions about differences between students and about statistical inferences that rely on those differences. This paper examines this problem in univariate IRT models. We will focus specifically on the three-parameter logistic (3PL) and the Rasch model.

In IRT, the latent variable is not directly observed but inferred from item responses through a statistical model. For the three-parameter logistic model, the probability of a correct response to item j is:

$$\pi_j(\theta) = P(y_j = 1 \mid \theta) = c_j + (1 - c_j)\frac{\exp(\alpha_j(\theta - \beta_j))}{1 + \exp(\alpha_j(\theta - \beta_j))}, \tag{1}$$

where θ denotes the person parameter (or latent trait), and β_j, α_j, and c_j represent the difficulty, discrimination, and guessing parameters of item j, respectively. Note that in the remainder of

[0]ORCID FT: http://orcid.org/0000-0002-1775-7654 ORCID: HR https://orcid.org/0000-0001-6290-8792 This paper is the result of follow-up work based on the first author's presidential address given at IMPS 2019 in Santiago de Chile. Thanks to David Torres Irribara, the associate editor, and the reviewers for constructive comments that have led to a better paper (all remaining errors are the responsibility of the authors). The code for generating the figures in this paper can be found on https://osf.io/62nqt. Contact information: francis.tuerlinckx@kuleuven.be. The research in this paper was supported in part by the Research Fund of KU Leuven (C14/23/062) and FWO Research project G050826N. AI was used as a language assistant in this paper and to give ideas on how to structure the paper.

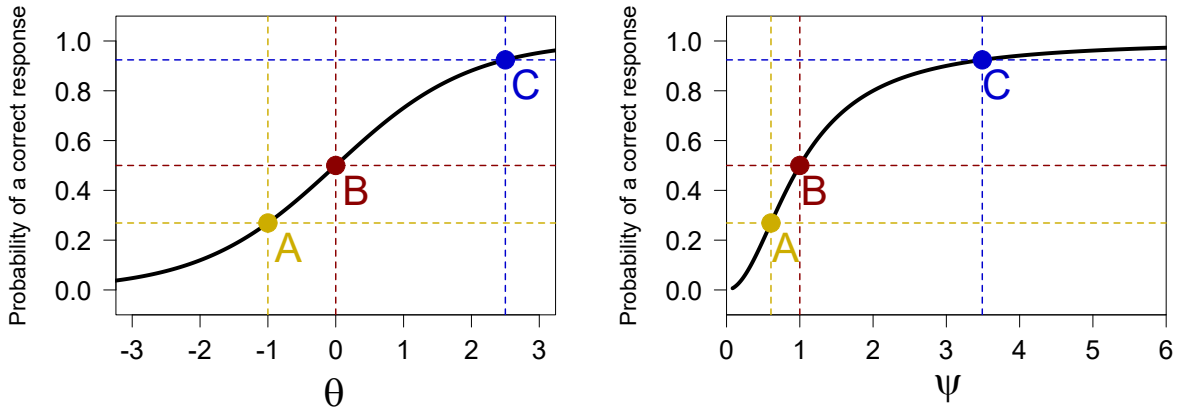

Figure 1: Two item characteristic curves for a Rasch item (i.e., $\alpha = 1, \beta = 0, c = 0$) under two different latent trait parametrizations. In the left panel, the latent trait or person parameter is θ (as in Equation 1). In the right panel, the person parameter is $\psi = e^{\frac{\theta}{2}}$. To illustrate the effect of the transformation clearly, we have selected three hypothetical persons (A, B, and C). It can be seen that for both parametrizations, the probability of correctly responding to the item does not change, but the relative positions of the persons on the latent trait do change.

this paper, we will assume that the item parameters (i.e., the β_j's, α_j's and c_j's) are known (e.g., from pretesting or calibration). The Rasch model is a special case of the 3PL model with $c_j = 0$ and $\alpha_j = 1$ for all items j.

Under smooth monotone transformations of the latent trait θ, the same response probabilities can be represented on a different numerical scale. Let $\psi = f(\theta)$ be a smooth strictly monotone transformation. Because the transformation is strictly monotone, its inverse exists and is also smooth and strictly monotone: $\theta = f^{-1}(\psi)$. In the remainder of this paper, we will suppress the function symbols f and f^{-1} and write in shorthand the transformation and its inverse as: $\psi = \psi(\theta)$ and $\theta = \theta(\psi)$.

Inserting the inverse transformation $\theta = \theta(\psi)$ into Equation 1 makes the probabilities dependent on another person parameter or latent trait (i.e., ψ), but does not change the response probabilities (see also, e.g., Feuerstahler, 2019; Lord, 1975). Using θ or ψ only represents the response probabilities of persons in a different coordinate system (i.e., a different parametrization). Although the person ordering is preserved (because of the strict monotonicity), the distances between persons are generally altered. This is illustrated graphically in Figure 1 for the Rasch model with $\psi = e^{\frac{\theta}{2}}$ (and thus the inverse transformation $\theta = 2\log(\psi)$).

This scale indeterminacy problem (i.e., the lack of a unique scale for the latent trait) has long been recognized in IRT (see e.g., Ballou, 2009; Feuerstahler, 2019; Lord, 1975; Ramsay,

1996). However, the issue has largely been ignored by relying on the θ scale by convention in most IRT analyses, publications, and software (a "measurement by fiat" approach).

But ignoring a problem does not resolve it and it matters because latent trait scores or person parameter estimates from IRT models are used in secondary analyses that compare groups, evaluate interventions, or support policy decisions. In such contexts, researchers often proceed with the conventional θ scale and rely on parametrization-dependent inferences. The problem does not lie with the IRT model itself, because a given model may be a very good model for the data, but with downstream uses of the person parameter estimates that are not invariant under reparametrizations[1].

The first (to the best of our knowledge) to address the scale indeterminacy problem using a geometric perspective is Ramsay (1996). Because the response probabilities under an IRT model are parametrization-invariant, Ramsay (1996) proposes to use these probabilities to define a parametrization-invariant ability measure. Specifically, Ramsay (1996) starts from the test performance curve (TPC; see also Ramsay & Wiberg, 2017), which is the curve defined by the probabilities of correctly solving each test item in a Euclidean space. Thus, if there are m items, then the TPC traces in the m-dimensional unit hypercube a curve defined as $\text{TPC}(\theta) = (\pi_1(\theta), \ldots, \pi_j(\theta), \ldots, \pi_m(\theta))$. An example with two Rasch items can be seen in Figure 2. Using the TPC, the distance between any two test takers (out of A, B, and C) can be calculated as the length of the path connecting them.

The approach of Ramsay (1996) leads to an alternative ability measurement for a particular person P, called here arc length ability and denoted as $L(P)$. The arc length represents the path length or distance along the TPC starting from the origin (where the probability of solving each of the items correctly is zero) to the actual position of person P on the curve. For the 3PL model, the arc length is calculated as follows for a person P with person parameter θ_P:

$$L(P) = \int_{-\infty}^{\theta_P} \sqrt{\sum_{j=1}^{m} \frac{\alpha_j^2 (1-c_j)^2 e^{2\alpha_j(\theta'-\beta_j)}}{\left(1+e^{\alpha_j(\theta'-\beta_j)}\right)^4}} d\theta'. \tag{2}$$

[1]The admissible transformations do not affect the ordering of individuals, but they do change the distances between them. The consequences of using numerical parameter values can be severe, because certain transformations may produce extreme values. One example is the transformation $\theta \mapsto \theta e^{\theta^2}$, which maps even moderately large absolute values of θ to very large numbers. If the original θ-values follow a standard normal distribution, the transformed values will be far from normally distributed. This will, of course, affect summary statistics, regression models, and other analyses that depend on the numerical values of the estimates.

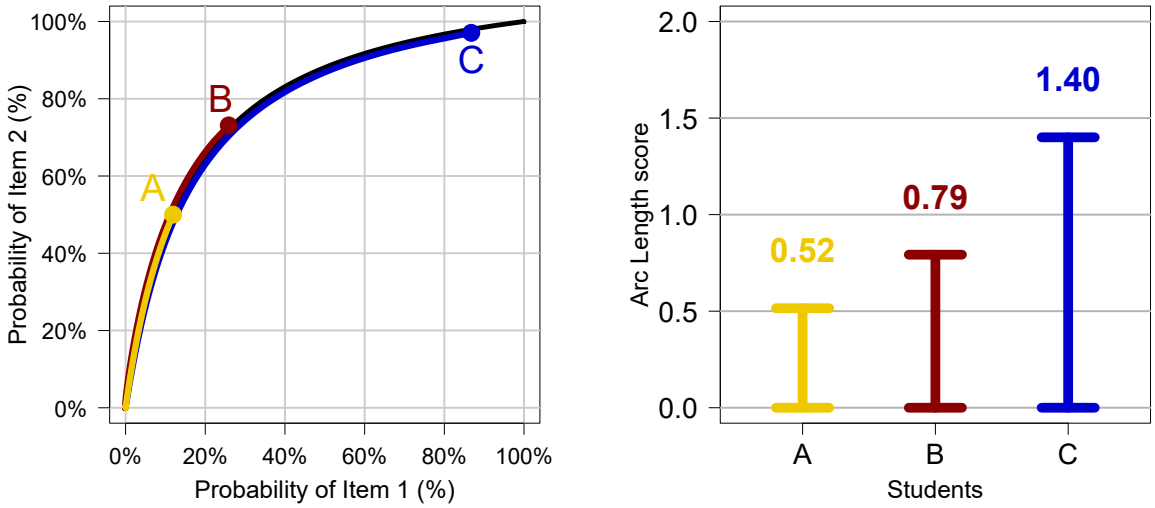

Figure 2: The left panel displays the test performance curve (TPC) in black for a two-item test of Rasch items. The colored lines depict the arc length scores for three students (A, B, and C). Arc length is measured as the length of the curve from its origin to the student's position. The right panel presents each student's computed arc length scores to facilitate the comparison.

Ramsay's arc length $L(P)$ is invariant under monotone transformations of the person parameter because it is based entirely on quantities that are invariant under monotone transformations themselves (i.e., the probabilities of solving items correctly).

Despite its conceptual appeal, the arc length has received little attention in psychometric research over the past few decades (but see e.g., Rossi, Wang, & Ramsay, 2002; Sijtsma & Meijer, 2006; Feuerstahler, 2020, 2023). The reasons for this relative neglect are unclear. One possible explanation is that arc length relies on differential geometry, a branch of mathematics that is not commonly associated with psychometrics and with which many researchers may be unfamiliar. At the same time, the concept may also appear somewhat ad hoc, insofar as it defines a measure by tracking distance along a probability curve. Together, these features may have limited its broader adoption[2].

In this paper, we want to further develop the arc length approach by providing a firm conceptual foundation. This will require the use of concepts from non-Euclidean differential geometry rather than from classical Euclidean differential geometry. The reason for this shift becomes apparent when reconsidering how distances based on arc length are actually computed. In the approach of Ramsay (1996), the distance between two persons (e.g., A and B in Figure 2) is

[2]Ramsay himself commented on this lack of attention of his paper stating in an interview that "it was perceived as another Dalí-esque intervention" (Genest & Nešlehová, 2014, p.177). In the same interview, Ramsay also remarked that "we should be using differential geometry much more than we do currently" (p.177)

defined as the length of the segment of the TPC connecting their locations, an idea that is intuitively appealing. However, for a specific statistical model (e.g., the Rasch or 3PL model), the TPC can be seen as a family of probability distributions $p(y \mid \theta)$, where $y = (y_1, \ldots, y_j, \ldots, y_m)$ denotes the binary response vector and θ the ability parameter (assuming known item parameters). Consequently, the locations of two persons on the TPC correspond to two distinct model instances, $p(y \mid \theta_A)$ and $p(y \mid \theta_B)$. Stated differently, $p(y \mid \theta_A)$ and $p(y \mid \theta_B)$ are two different points in a model space. As shown by Rao (1945) and elaborated in Amari (2016) and Calin and Udrişte (2014), these model instances generally do not just form a curve in Euclidean space as is (implicitly) assumed by Ramsay, 1996). Rather, these model instances form a curved, non-Euclidean space. (Note that the field of non-Euclidean differential geometrical analysis of statistical models is also called information geometry; (Amari, 2016).)

This paper makes two contributions. First, we will derive a parametrization-invariant distance for the 3PL and Rasch models, which we call *geodesic ability*. Crucial in this distance calculation is the Fisher information, which is used to define the Fisher-Rao metric, and then for measuring how different two nearby persons are in terms of their probability distributions implied by the model. Second, we compare the new geodesic ability to Ramsay's arc-length proposal. Both proposals are motivated by the same concerns but they arise from different geometric structures (Euclidean for Ramsay's arc length and non-Euclidean for the geodesic ability). As a result, the arc-length and geodesic ability will be similar, but not equal.

The remainder of the paper is organized as follows. First, we start by introducing the differential-geometric framework, followed by a derivation of the geodesic ability for the 3PL model. Next, we study the special case of the Rasch model in more detail. We end the paper with a conclusion.

2 Non-Euclidean differential geometry

In this section, we will explain briefly the basic notions of differential geometry for statistical models. Interested readers can consult various sources for more information (see e.g., Amari, 2016; Boothby, 1986; Tu, 2010; Lee, 2012). The relation between statistics and differential geometry is more deeply elaborated in Amari (2016), Calin and Udrişte (2014), Kass (1989), and Kass and Vos (2011). Two noteworthy papers from quantitative psychology with nice

introductions into some of the material used in this paper are Ly, Marsman, Verhagen, Grasman, and Wagenmakers (2017) and Segert and Davis-Stober (2019).

A parametric statistical model can be viewed as a collection $\mathcal{M}$ of probability distributions[3][4] (with outcome space $\mathcal{S}$) indexed by a parameter θ: $\mathcal{M} = \{p(y|\theta) \,|\, \theta \in \Theta\}$ (McCullagh, 2002). The parameter θ can be a scalar (as in the 3PL model from Equation 1) or, more generally, a vector of dimension k. The set $\Theta \subseteq \mathbb{R}^k$ is called the parameter space. The tools in this paper require that we are dealing with well-behaving statistical models. By well-behaving, we mean that the model should be well-identified (i.e., $\theta \neq \theta'$ implies that $p(y|\theta) \neq p(y|\theta')$) and satisfy general regularity conditions. In addition, if the model is multidimensional, then the Fisher information matrix should be of full rank for all θ. The most common statistical models belong to this class of regular statistical models (e.g., normals with unknown location and scale, exponential family models, etc.).

A regular parametric statistical model $\mathcal{M}$ constitutes a smooth (or differentiable) manifold of distributions, which means that an (infinitely) differentiable and invertible map exists from the manifold to the parameter space $\Theta \subseteq \mathbb{R}^k$ (such a map is called a chart map or chart). The dimensionality k of the map is the dimensionality of the manifold. For instance, a two-dimensional statistical manifold is illustrated graphically in Figure 3. A smooth manifold of regular statistical distributions is called here a *statistical manifold*.

A simple example is the family of normal distributions $\mathcal{M} = \{p(y|\mu, \sigma) = N(\mu, \sigma)|(\mu, \sigma) \in \Theta\}$, $\Theta = \mathbb{R} \times (0, \infty)$. Here, each pair (μ, σ) determines one probability distribution, so the model is indexed by two parameters and therefore has dimension $k = 2$. The mapping that assigns to each distribution in $\mathcal{M}$ its parameter values (μ, σ) is a chart, and it provides a coordinate system for the manifold. Since the density of the normal distribution depends smoothly on μ and σ (all derivatives of $p(y|\mu, \sigma)$ with respect to μ and σ are continuous), this family forms a smooth statistical manifold. In this case, the manifold can be understood as the collection of all normal distributions, while (μ, σ) gives one particular coordinate system for describing them.

Different parametrizations of the model provide different coordinate systems on the same

[3]We will use the shorthand term "distributions" referring to either a "probability mass functions" or a "probability density function", depending on the context.

[4]Starting from a more general definition, a statistical model is a family of probability measures on a measurable space. In this work, we restrict attention to models whose distributions admit densities (probability mass functions or probability density functions) with respect to a counting or Lebesgue measure.

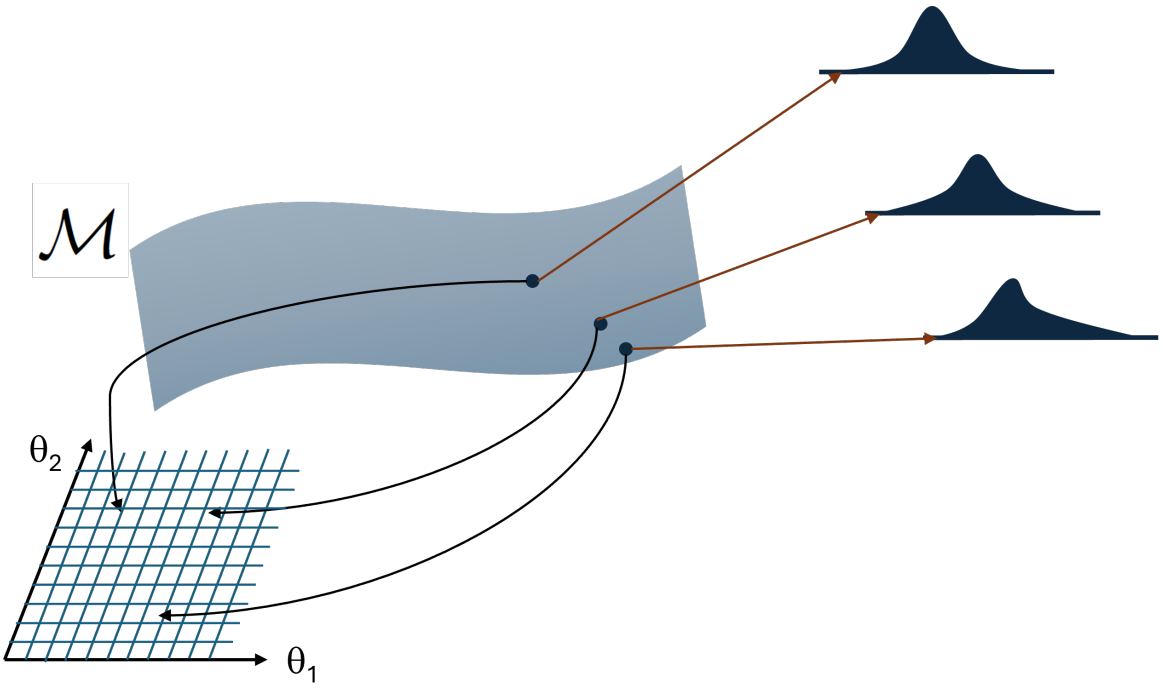

Figure 3: Graphical illustration of a smooth two-dimensional statistical manifold $\mathcal{M}$ together with a coordinate or parameter space Θ. Each location on the manifold corresponds to a distribution of some family and (θ_1, θ_2) is a pair of coordinates indexing the distributions. The mapping between $\mathcal{M}$ and the parameter space is represented by the black lines.

statistical manifold. For instance, consider the statistical manifold for the univariate normal distribution, which can be parametrized with (μ, σ) (mean and standard deviation), but also with (μ, σ^2) (mean and variance), or with $(\mu, 1/\sigma^2)$ (mean and precision, as is common in Bayesian statistics). These different parametrizations provide different coordinate systems to distinguish the distributions of the normal statistical manifold, in the same way as there exist different types of 2D geographical maps that distinguish between actual places on the earth. This comparison also makes clear that the manifold (be it the earth or a statistical manifold) and its properties (e.g., curvature, distance between objects) does not depend on the coordinate system or parametrization that is used.

The statistical manifold of the 3PL model with known item parameters from Equation 1 is a one-dimensional manifold. The most common parametrization is based on $\theta \in \Theta = \mathbb{R}$. However, as explained in the introduction, any smooth strictly monotone transformation of θ can be used as well: $\psi = e^{\frac{\theta}{2}}$, $\omega = \sqrt[3]{\theta}$, etc.

A particular choice of parametrization of the statistical manifold has to be made in order to perform calculations on the statistical manifold. For example, finding the best fitting model (i.e., the point on the statistical manifold best representing the data) requires a par-

ticular parametrization that allows for parameter estimation. However, the arbitrariness of the parametrization requires a researcher to think about which quantities are invariant under reparametrization and which are not. As shown in Figure 1, the difference between two persons' θ scores is not parametrization-invariant. Thus we seek to derive a notion of distance between two persons that is intrinsic to the model's geometric structure rather than depending on the chosen coordinate system or parametrization.

A first step is to find an expression for an infinitesimal distance between probability distributions. Consider a parametric family of probability densities $p(y \mid \theta)$, where $\theta \in \Theta$ (with dimension k), and two nearby points θ and $\theta + d\theta$ in parameter space. In the Euclidean parameter space, the squared distance between these points is given by

$$\delta^2(\theta, \theta + d\theta) = \sum_i d\theta_i^2 = d\theta^\top d\theta.$$

This expression, however, uses the notion of a distance in a Euclidean space and it does not reflect how easily one can distinguish between different probability distributions in the close vicinity of θ (for an elaboration on this, see e.g., Myung, Balasubramanian, & Pitt, 2000). Taking this into account (see the derivation in Appendix A), we see that the squared infinitesimal distance between two nearby distributions becomes:

$$(ds)^2 = d\theta^\top g(\theta)\, d\theta, \tag{3}$$

where $g(\theta)$ is the Fisher information matrix. This squared distance is called the squared line element, or also the Fisher-Rao metric. Equation 3 can be considered a generalization of the well-known Pythagorean theorem for non-Euclidean geometry. In general, $(ds)^2$ is not equal to the squared distance in the (Euclidean) parameter space (i.e., $\delta^2(\theta, \theta+d\theta)$), unless the statistical manifold itself is a "flat" Euclidean space, in which case $g(\theta) = I_k$ (the $k \times k$ identity matrix) for all θ. In the univariate case, the Fisher-Rao metric is defined as:

$$(ds)^2 = g(\theta)\, (d\theta)^2. \tag{4}$$

Remark. The squared line element $(ds)^2$ informs us about the local distinguishability between $p(y|\theta)$ and its nearby distributions. A more common quantity to assess differences between

distributions is the Kullback-Leibler divergence D_{KL}. Using D_{KL}, the difference of $p(y|\theta')$ from $p(y|\theta)$ is:

$$D_{KL}\left(p(y|\theta)\|p(y|\theta')\right) = \int_S p(y|\theta)\log\left(\frac{p(y|\theta)}{p(y|\theta')}\right)dy.$$

Although $D_{KL} \geq 0$ (with equality when $p(y|\theta) = p(y|\theta')$ for all y), it is not symmetric, nor satisfies the triangle inequality, and thus the D_{KL} divergence is not a distance. However, when $\theta' = \theta + d\theta$, the quadratic approximation to D_{KL} results in:

$$D_{KL}\left(p(y|\theta)\|p(y|\theta + d\theta)\right) \approx \frac{1}{2}(ds)^2,$$

which means that for nearby distributions, the D_{KL} divergence is actually carrying the same information as the local distinguishability $(ds)^2$. (This ends the remark.)

Returning to the main argument, the squared line element $(ds)^2$ is called the Fisher-Rao metric because Rao (1945) was the first to show that Fisher's information matrix $g(\theta)$ is in fact a so-called metric tensor[5] that is used to measure distances on a smooth manifold equipped with this metric. A differentiable manifold endowed with such a metric is called a Riemannian manifold in differential geometry. An important property of $(ds)^2$ is stated in the proposition below.

Proposition 1. *The Fisher-Rao metric is invariant under smooth reparametrizations.*

Proof. We will prove this for the univariate case. For a general proof for the parameter vector case, we refer to Calin and Udrişte (2014).[6]

Assume that $\phi = \phi(\theta)$ is a smooth reparameterization. This implies that $\phi(\cdot)$ is differentiable, one-to-one and invertible. The inverse transformation is $\theta = \theta(\phi)$.

Let us denote the distribution under the new ϕ-parametrization as $p^*(y|\phi) = p(y|\theta(\phi))$ and the Fisher information as $g^*(\phi)$. Then we have to show that $(ds)^2 = (ds^*)^2$.

Because $g^*(\phi) = E\left[\left(\frac{d\log p^*(y|\phi)}{d\phi}\right)^2\right]$, we need an expression for $\frac{d}{d\phi}\log p^*(y|\phi)$. Using the chain rule, we find that $\frac{d\log p^*(y|\phi)}{d\phi} = \frac{d\log p(y|\theta(\phi))}{d\phi} = \left.\frac{d\log p(y|\theta)}{d\theta}\right|_{\theta=\theta(\phi)}\frac{d\theta}{d\phi}$.

Inserting this in the definition of the Fisher information gives (we write an integral over the

[5]More exactly, it is a (0,2) metric tensor, also called 2-covariant tensor.

[6]To be more precise: Calin and Udrişte (2014) prove that the Fisher information matrix is covariant under smooth reparametrization, but that cancels against the contravariant change in $d\theta$ under reparametrization.

sample space $\mathcal{S}$, but if $\mathcal{S}$ is discrete, the integral should be understood as a summation):

$$\begin{aligned} g^*(\phi) &= \int_{\mathcal{S}} p^*(y|\phi) \left(\frac{d \log p^*(y|\phi)}{d\phi} \right)^2 dy \\ &= \int_{\mathcal{S}} p(y|\theta(\phi)) \left(\left. \frac{d \log p(y|\theta)}{d\theta} \right|_{\theta=\theta(\phi)} \frac{d\theta(\phi)}{d\phi} \right)^2 dy \\ &= \int_{\mathcal{S}} p(y|\theta(\phi)) \left(\left. \frac{d \log p(y|\theta)}{d\theta} \right|_{\theta=\theta(\phi)} \right)^2 dy \quad \left(\frac{d\theta(\phi)}{d\phi} \right)^2 \\ &= g\big(\theta(\phi)\big) \left(\frac{d\theta(\phi)}{d\phi} \right)^2, \end{aligned} \tag{5}$$

or in shorter (but somewhat sloppier) notation: $g^*(\phi) = g(\theta) \left(\frac{d\theta}{d\phi} \right)^2$. Because the differential $d\phi$ transforms as well (but in the opposite way), $d\phi = \frac{d\phi}{d\theta} d\theta$, the squared line element $(ds^*)^2$ becomes:

$$(ds^*)^2 = g^*(\phi)\,(d\phi)^2 = g(\theta) \left(\frac{d\theta}{d\phi} \right)^2 \left(\frac{d\phi}{d\theta} \right)^2 (d\theta)^2 = g(\theta)\,(d\theta)^2 = (ds)^2,$$

so the (squared) line element (i.e., the Fisher-Rao metric) does not change. □

In the multivariate case, the result generalizes to: $(ds^*)^2 = d\phi^\top\, g^*(\phi)\, d\phi = d\theta^\top\, g(\theta)\, d\theta = (ds)^2$.

Remark. The line element ds is a dimensionless quantity. Let us start with the scalar parameter θ. As is well-known from statistics, the inverse of Fisher information $g(\theta)$ is the variance of an estimator for the parameter θ. Therefore, $[g(\theta)] = [\theta]^{-2}$, that is, the dimension of the Fisher information is equal to the dimension of θ^{-2}. Because $(d\theta)^2$ has dimension of θ^2, it follows that $[d\theta]^2 = [\theta]^2$. Combining the ingredients gives: $[ds]^2 = [g(\theta)][d\theta]^2 = [\theta]^{-2}[\theta]^2 = 1$. Thus, $(ds)^2$ is dimensionless and therefore also ds is dimensionless. (This ends the remark.)

A distance on the manifold $\mathcal{M}$ between two arbitrary (not necessarily nearby) distributions $p_a = p(y|\theta_a)$ and $p_b = p(y|\theta_b)$ can then be calculated by defining a path between them and then computing the length of the path (i.e., an arc length on the statistical manifold). In general, the parameter θ has k components. Assume a smooth parametrized curve $\theta(t) = (\theta_1(t), \ldots, \theta_k(t))$ with $t \in [0, 1]$ such that $\theta(0) = \theta_a$ and $\theta(1) = \theta_b$. The arc length on the manifold $\mathcal{M}$ is found by assuming a fine-grained grid along the path and then accumulating the many small line

elements ds:

$$\begin{aligned} d_{\theta(t)}(p_a, p_b) &= d_{\theta(t)}(p(y|\theta_a), p(y|\theta_b)) \\ &= \int_{\text{along path } \theta(t)} ds \\ &= \int_0^1 \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k} g_{ij}(\theta(t)) \frac{d\theta_i(t)}{dt}\frac{d\theta_j(t)}{dt}}\, dt. \end{aligned} \tag{6}$$

Note that for the distance or arc length on the manifold, we use as a symbol[7] $d_{\theta(t)}$ to denote that in general the distance depends on the path taken by the curve $\theta(t)$. Another path $\tilde{\theta}(t)$ will likely lead to another arc length.

One particular important path is the one with the shortest possible arc length $d_{\min}$, called the geodesic distance:

$$d_{\min}(p_a, p_b) = \min_{\theta(t)} d_{\theta(t)}(p(y|\theta_a), p(y|\theta_b)).$$

The corresponding path is called the geodesic curve, denoted here as $\theta_{\min}(t)$:

$$\theta_{\min}(t) = \arg\min_{\theta(t)} d_{\theta(t)}(p_a, p_b).$$

The geodesic distance is invariant under smooth parametrizations. Moreover, it is a genuine distance (using p, q and r for distributions): (1) $d_{\min}(p, q) \geq 0$ with $d_{\min}(p, p) = 0$ (non-negativity), (2) $d_{\min}(p, q) = d_{\min}(q, p)$ (symmetry), and (3) $d_{\min}(p, q) + d_{\min}(q, r) \geq d_{\min}(p, r)$ (triangle inequality) (Lee, 2012).

Finding the geodesic curve is in most situations not straightforward and requires the solution to a variational problem (see e.g., Atkinson & Mitchell, 1981; Calin & Udrişte, 2014; Gelman & Meng, 1998). For a number of common distributions, these geodesic distances have been derived and can be found in various sources, (see e.g. Miyamoto, Meneghetti, Pinele, & Costa, 2024; Atkinson & Mitchell, 1981; Calin & Udrişte, 2014).

[7]The symbol d for distance should be not be confused with the differential that we used before.

3 Geodesic ability in the 3PL model

Let us now return to the 3PL model. Consider a single person (with unknown ability) taking a test of m items (and known item parameters α_j, β_j and c_j). The probability of a particular y_j response on a single item j is given by $\pi_j(\theta)^{y_j}(1-\pi_j(\theta))^{1-y_j}$ and the joint probability for the m items is:

$$p(y|\theta) = \prod_{j=1}^{m} \pi_j(\theta)^{y_j}(1-\pi_j(\theta))^{1-y_j} \tag{7}$$

where $\pi_j(\theta)$ is defined in Equation (1). As mentioned above, with known item parameters, the 3PL model is a one-dimensional statistical manifold and thus the Fisher information matrix is a scalar (Lord, 1980):

$$\begin{aligned} g(\theta) &= E\left[\left(\frac{d\log p(y|\theta)}{d\theta}\right)^2\right] \\ &= \sum_{j=1}^{m} \frac{\alpha_j^2(1-c_j)}{(c_j + e^{\alpha_j(\theta-\beta_j)})(1+e^{\alpha_j(\theta-\beta_j)})^2} \end{aligned}$$

In this context, the Fisher information is also called the test information function (because it gives the amount of information provided about θ by the particular test of m items).

Proposition 2. *For the one-dimensional 3PL model with known item parameters, the geodesic distance between two persons located at θ_P and θ_Q (with $\theta_Q < \theta_P$) is*

$$d(Q,P) = \int_{\theta_Q}^{\theta_P} \sqrt{g(\theta)}\, d\theta = \int_{\theta_Q}^{\theta_P} \sqrt{\sum_{j=1}^{m} \frac{\alpha_j^2(1-c_j)}{(c_j + e^{\alpha_j(\theta'-\beta_j)})(1+e^{\alpha_j(\theta'-\beta_j)})^2}}\, d\theta' \tag{8}$$

Proof. Consider Equation 6. On a one-dimensional statistical manifold with squared line element $(ds)^2 = g(\theta)\,(d\theta)^2$, the path as the image of the parametric curve $\theta(t)$ between $\theta_Q = \theta(0)$ and $\theta_P = \theta(1)$ has length

$$L = \int_0^1 \sqrt{g(\theta(t))}\left|\frac{d\theta(t)}{dt}\right| dt.$$

Because non-monotone curves will always lead to larger lengths compared to monotone curves, we will only consider the latter and thus $\frac{d\theta(t)}{dt} > 0$. The image of these monotone curves is the path connecting θ_Q and θ_P. Moreover, it is safe to state that for the statistical manifold of the 3PL model there is only one path connecting the points θ_Q and θ_P.

Now, we can do a change of variable $\theta' = \theta(t)$ (so that $d\theta' = \frac{d\theta(t)}{dt}dt$), the length is therefore

$$d(Q,P) = \int_0^1 \sqrt{g(\theta(t))}\,\frac{d\theta(t)}{dt}\,dt = \int_{\theta_Q}^{\theta_P} \sqrt{g(\theta')}\,d\theta'.$$

□

Remark. Following Atkinson and Mitchell (1981), an alternative derivation is possible in the one-dimensional situation: Find a transformation $\phi = \phi(\theta)$, such that the metric of ϕ is that of a Euclidean space (i.e., $(ds)^2 = (d\phi)^2$ with $g^*(\phi) = 1$). Such a map only exists in general for one-dimensional manifolds, but not for higher dimensional ones, which makes this method a "trick" that cannot be applied in general.

Thus, we are looking for a parameter ϕ such that

$$(ds)^2 = g^*(\phi)\,(d\phi)^2 = (d\phi)^2,$$

because $g^*(\phi) = 1$ for all ϕ. In that case, the geodesic distance between two test takers can be simply computed as $\phi_P - \phi_Q$. Expressing this difference as a function of the familiar θ then gives the solution.

To find the map from θ to ϕ (i.e., $\phi = \phi(\theta)$) so that the result is a Euclidean metric, we start with the requirement that $g^*(\phi) = g(\theta)\,\left(\frac{d\theta}{d\phi}\right)^2 = 1$. Then, taking the square root of both sides and solving for ϕ gives the result:

$$\phi = \int \sqrt{g(\theta')}\,d\theta' + C.$$

The integration constant C will cancel when taking differences between two ϕ values. (This ends the remark.)

Now that we have an expression for the geodesic distance between two persons P and Q (such that $\theta_Q < \theta_P$), we can also replace Q with a reference person that has the lowest possible trait score (i.e., $\theta_Q \to -\infty$). Using Equation (8), gives a corresponding distance defined relative to a fixed origin at $-\infty$:

$$A(P) = \int_{-\infty}^{\theta_P} \sqrt{g(\theta')}\,d\theta' = \int_{-\infty}^{\theta_P} \sqrt{\sum_{j=1}^{m} \frac{\alpha_j^2(1-c_j)}{(c_j + e^{\alpha_j(\theta'-\beta_j)})(1+e^{\alpha_j(\theta'-\beta_j)})^2}}\,d\theta' \tag{9}$$

We call $A(P)$ the *geodesic ability* of person P for the given test. To the best of our knowledge, there exists no closed-form solution to this integral.

The interpretation of $A(P)$ differs from that of the conventional latent trait value θ_P. Equal increments in θ do not in general correspond to equal statistical separation between persons, because the local geometry varies with $g(\theta)$. By contrast, equal increments in geodesic ability correspond to an equal distance along the 3PL model manifold.

Proposition 3. *The geodesic ability A is invariant under smooth reparametrizations of the latent trait.*

Proof. Let $\phi = \phi(\theta)$ be a smooth transformation. Using Equation 5, the Fisher information expressed as a function of ϕ satisfies:

$$g^*(\phi) = g(\theta(\phi)) \left(\frac{d\theta}{d\phi}\right)^2.$$

Therefore,

$$\int \sqrt{g^*(\phi)}\, d\phi = \int \sqrt{g(\theta(\phi))} \left|\frac{d\theta}{d\phi}\right| d\phi = \int \sqrt{g(\theta(\phi))} \frac{d\theta}{d\phi}\, d\phi = \int \sqrt{g(\theta)}\, d\theta,$$

where the absolute value has been removed because the transformation is monotone. This shows that the distance is unchanged. □

Comparing the geodesic ability $A(P)$ (Equation 9) and Ramsay's arc length $L(P)$ (Equation 6), it can be concluded that the equations are quite similar, but not the same. The difference comes from the fact that in our derivation, we have taken into account the metric of the statistical manifold. However, more detailed comparisons can be made for the Rasch model.

4 Geodesic ability in the Rasch Model

We will now consider the special case of the Rasch model ($\alpha_j = 1, c_j = 0$ for all j). The geodesic ability becomes:

$$A(P) = \int_{-\infty}^{\theta_P} \sqrt{\sum_{j=1}^{m} \frac{e^{\theta' - \beta_j}}{\left(1 + e^{\theta' - \beta_j)}\right)^2}} d\theta', \tag{10}$$

This can be compared to Ramsay's arc length for the Rasch model:

$$L(P) = \int_{-\infty}^{\theta_P} \sqrt{\sum_{j=1}^{m} \left[\frac{e^{\theta' - \beta_j}}{\left(1 + e^{\theta' - \beta_j)}\right)^2} \right]^2} d\theta'.$$

The formulas are even more similar than for the 3PL model. The only difference is that the terms of the sum under the radical in Ramsay's arc length are squared.

As for the 3PL model, there exists no closed-form expression for the geodesic ability in the Rasch model. However, we can get more insight by considering a further simplification: Assume that the m items all have difficulty equal to 0. In this case, the geodesic distance between two persons P and Q reduces to (using the subscript 0 to refer to the special case):

$$\begin{aligned} d_0(Q, P) &= \sqrt{m} \int_{\theta_Q}^{\theta_P} \sqrt{\frac{e^{\theta'}}{\left(1 + e^{\theta'}\right)^2}} d\theta' \\ &= 2\sqrt{m} \left(\tan^{-1}\left(e^{\frac{\theta_P}{2}}\right) - \tan^{-1}\left(e^{\frac{\theta_Q}{2}}\right) \right) \end{aligned}$$

Note that this result is actually equal to the geodesic distance between two binomial distributions as derived by Atkinson and Mitchell (1981). This should not come as a surprise because in the equal item difficulty case the Rasch model with m items reduces to a binomial model (because the success probability is constant).

Similarly to equation (9), we can set $\theta_Q \to -\infty$ to obtain the geodesic ability for person P, denoted now with a subscript 0 as $A_0(P)$. Thus, it simplifies to:

$$A_0(P) = 2\sqrt{m} \tan^{-1}\left(e^{\frac{\theta_P}{2}}\right).$$

It can be seen that $\lim_{\theta_P \to -\infty} A_0(P) = 0$ and $\lim_{\theta_P \to +\infty} A_0(P) = \pi\sqrt{m}$.

Figure 4 graphically illustrates $A(P)$ and $A_0(P)$ as functions of θ_P for various choices of β and number of items m. Several observations can be made from this plot. First, the plots confirm that the geodesic ability starts at 0 and is positive. Second, the major factor that determines where the function levels off is the number of items. Third, if the items are placed symmetrically around zero, $A_0(P)$ serves as a decent approximation to $A(P)$.

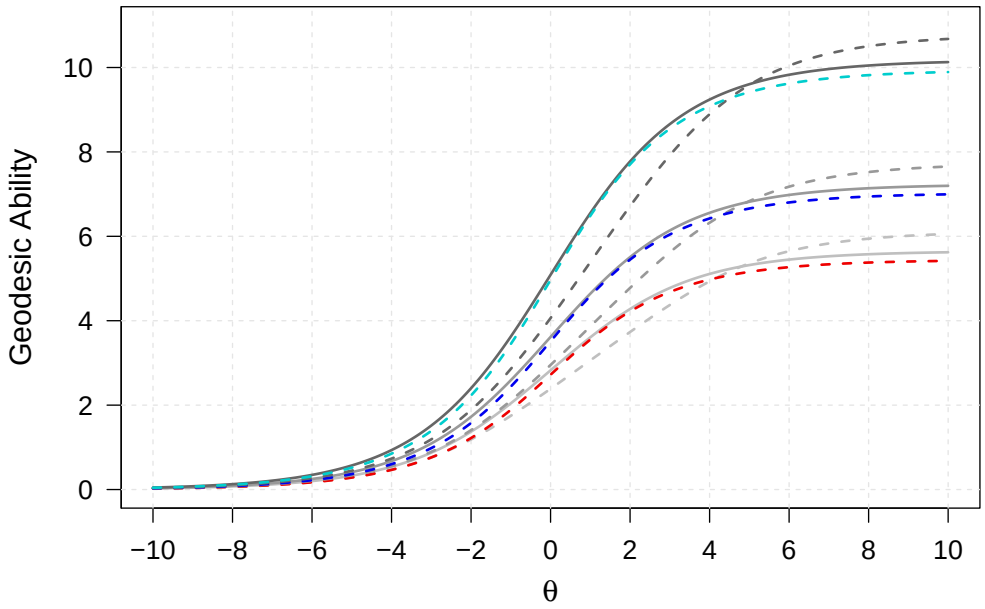

Figure 4: The geodesic ability $A(P)$ (or $A_0(P)$) as a function of θ_P. The sets of lines with the same color correspond to the same test length: $m = 3$ (red), $m = 5$ (blue) and $m = 10$ (cyan). The dotted lines stand for $A_0(P)$ (thus all item difficulties equal to zero). The solid and dashed lines correspond to tests with item difficulties located symmetrically or asymmetrically around zero, respectively. More specifically, for the solid lines from bottom to top, the following β values (denoted as $a(b)c$, such that a sequence starts at a, ends at c and has b units between consecutive values) are being used: $(-1(1)1)$, $(-1(0.5)1)$, $(-1(2/9)1)$. The dashed lines have asymmetrically placed β values: $(-1(2)3)$, $(-1(1)3)$, $(-1(4/9)3)$.

Suppose one has to estimate the geodesic ability of a person based on a response vector and known (or calibrated) item difficulties. The standard IRT software can be used to obtain an estimator of θ, denoted as $\hat{\theta}$. Plugging this estimate into Equation 10 gives the estimated geodesic ability: $\widehat{A(P)}$. It is also important to quantify the uncertainty of this estimator. To study this, let us assume again a test of m items with all $\beta_j = 0$. Thus, we want to assess the uncertainty of $\widehat{A_0(P)}$. From statistical theory, we know that $\sqrt{m}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, m \cdot g^{-1}(\theta)\right)$, where $g^{-1}(\theta)$ is the reciprocal of the Fisher information (or, in the multiparameter case, it is the inverse of the Fisher information matrix). Next, we apply the delta method by first computing

$A_0'(P) = \frac{dA_0(P)}{d\theta} = \sqrt{m}\frac{e^{\frac{\theta_P}{2}}}{1+e^{\theta_P}}$. This then gives:

$$\sqrt{m}\left(\widehat{A_0(P)} - A_0(P)\right) \xrightarrow{D} N\left(0, mg^{-1}(\theta)\left(A_0'(P)\right)^2\right)$$
$$\xrightarrow{D} N(0,1),$$

because if all $\beta_j = 0$, then $g(\theta) = m\frac{e^{\theta}}{\left(1+e^{\theta}\right)^2}$. Consequently, $SE\left(\widehat{A_0(P)}\right) \approx 1$. Hence, we have an estimator with a constant uncertainty. This will not hold exactly for the general $\widehat{A(P)}$, but it can be expected to hold approximately.

It is interesting to note that this result coincides with the findings by Hougaard (1982). Based on earlier work, Hougaard (1982) studied alternative parametrizations with certain properties. More specifically, for an exponential family model[8], the following class of transformations of the canonical parameter θ is considered:

$$B_\delta(\theta) = \int_L^\theta \left[\frac{d^2}{d\theta'^2}\Psi(\theta')\right]^\delta d\theta',$$

where L is the lower bound for the parameter θ (in our case, $L = -\infty$) and δ is a constant. The choice of δ determines a new parametrization with a particular property. For example, when $\delta = \frac{1}{3}$, this results in a normal likelihood. The $\delta = \frac{1}{2}$ leads to a parametrization with stability of variance. In fact, for $\delta = \frac{1}{2}$, it becomes clear that $B_{\frac{1}{2}}(\theta) = A(P)$ for the Rasch model. Moreover, we have shown the constant variance for the geodesic ability.

5 Conclusion

In this paper we addressed the problem of scale indeterminacy in unidimensional IRT models with known item parameters. Under monotone transformations of the latent trait, the probabilities remain the same, but the scores assigned to persons may be very different. This casts doubt on the use of a person parameter to characterize the ability or latent trait value of an individual.

Using differential geometry, we showed that the Fisher-Rao metric induces an intrinsic distance on the one-dimensional IRT manifold and that this distance gives rise to a corresponding

[8]The main focus of Hougaard (1982) is on the curved exponential family model.

score, geodesic ability $A(P)$. This score is invariant under smooth monotone reparametrizations. Because of this invariance property, the geodesic ability provides a principled alternative to arbitrary latent variables.

The geodesic ability $A(P)$ shares some important properties and similarities with Ramsay's arc length $L(P)$, which has been motivated on the same grounds as the geodesic ability. First, both quantities are derived using a geometrical argument. Second, both ability measures remain invariant under monotone transformations of the latent variable. At the same time, Ramsay's construction and geodesic ability are not identical. Ramsay's arc length is defined for a curve embedded in a Euclidean space. Geodesic ability, by contrast, is defined on the statistical manifold of the IRT model using the Fisher-Rao metric. The statistical manifold of the IRT model is a curved space and the geodesic ability takes this curvature into account through the Fisher information. In fact, one could say that the geodesic ability is an intrinsic distance induced by the model itself. Ramsay's arc length is defined using the test performance curve. This is a curve in a k-dimensional (Euclidean) space, which acts as an ambient space for the curve. Such an ambient space is not needed in the case of the geodesic ability. The only information needed to calculate a distance is the Fisher information.

The geodesic ability proposed in this paper will need some further investigation. One issue that needs to be elaborated is the generalization to multiple dimensions and the corresponding interpretation. Another aspect that deserves further study is the fact that the geodesic ability depends on m, the number of items in the test. A straightforward way of dealing with this problem is to define a normalized version of $A(P)$: $\frac{A(P)}{m}$. This kind normalization is most likely too crude to work well (since the upper limit of $A(P)$ depends on the specific item difficulties), but it may be worthwhile to elaborate further.

The first results obtained from a differential-geometrical analysis for simple unidimensional IRT offer a potential useful way of reconsidering old problems in IRT and latent variable modeling.

Appendices

A Derivation of $(ds)^2 = d\theta^\top g(\theta)\, d\theta$ (Equation 3)

In this Appendix, we will derive an expression for an infinitesimal distance between probability distributions. For reasons of generality, we will focus on the multivariate case, that is, Equation 3. The derivation follows the original argument of Rao (1945) (see also Caticha, 2022).

To compare the distributions $p(y \mid \theta)$ and $p(y \mid \theta+d\theta)$ directly, define their relative difference

$$\Omega(y) = \frac{p(y \mid \theta + d\theta) - p(y \mid \theta)}{p(y \mid \theta)}. \tag{11}$$

Using a first-order Taylor expansion,

$$p(y \mid \theta + d\theta) \approx p(y \mid \theta) + \sum_i \frac{\partial p(y \mid \theta)}{\partial \theta_i}\, d\theta_i,$$

we obtain

$$\Omega(y) \approx \sum_i \frac{1}{p(y \mid \theta)} \frac{\partial p(y \mid \theta)}{\partial \theta_i}\, d\theta_i = \sum_i \frac{\partial \log p(y \mid \theta)}{\partial \theta_i}\, d\theta_i.$$

Taking the expectation of $\Omega(y)$ with respect to $p(y \mid \theta)$ yields

$$E[\Omega(y)] = 0,$$

so $\Omega(y)$ itself is unsuitable as a distance measure. A meaningful quantity is obtained instead from the expected squared relative difference, which represents the mean squared relative de-

viation, also called distinguishability, between the two nearby distributions:

$$
\begin{aligned}
E\left[\Omega(y)^2\right] &\approx \int_S \left(\frac{1}{p(y|\theta)} \sum_i \frac{\partial p(y|\theta)}{\partial \theta_i} d\theta_i\right)^2 p(y|\theta) dy \\
&= \int_S \left(\frac{1}{p(y|\theta)} \sum_i \frac{\partial p(y|\theta)}{\partial \theta_i} d\theta_i\right) \left(\frac{1}{p(y|\theta)} \sum_j \frac{\partial p(y|\theta)}{\partial \theta_j} d\theta_j\right) p(y|\theta) dy \\
&= \int_S \left(\sum_i \frac{\partial \log p(y|\theta)}{\partial \theta_i} d\theta_i\right) \left(\sum_j \frac{\partial \log p(y|\theta)}{\partial \theta_j} d\theta_j\right) p(y|\theta) dy \\
&= \sum_{i,j} E\left[\left(\frac{\partial \log p(y|\theta)}{\partial \theta_i}\right) \left(\frac{\partial \log p(y|\theta)}{\partial \theta_j}\right)\right] d\theta_i d\theta_j \\
&= ds^2, \qquad (12)
\end{aligned}
$$

The expected value of the product of first derivatives of the log density in Equation 12 is the (i, j)th element of the Fisher information matrix $g(\theta)$. In matrix notation:

$$
g(\theta) = E\left[\frac{\partial \log p(y|\theta)}{\partial \theta^\top} \frac{\partial \log p(y|\theta)}{\partial \theta}\right].
$$

For the statistical models considered in this paper, it can be shown $g_{ij}(\theta) = -E\left[\frac{\partial^2 \log p(y|\theta)}{\partial \theta_i \partial \theta_j}\right]$ (see Calin & Udrişte, 2014).

The resulting squared line element on the statistical manifold is therefore

$$
(ds)^2 = \sum_{i,j} g_{ij}(\theta) d\theta_i d\theta_j = d\theta^\top g(\theta) d\theta. \qquad (13)
$$

This expression generalizes the Euclidean Pythagorean theorem. Only when $g(\theta) = I_k$ for all θ does $(ds)^2$ reduce to $\delta^2(\theta, \theta + d\theta)$. The derivation also clarifies why the squared relative difference provides a natural starting point for defining distance on a statistical manifold.

References

Amari, S. (2016). *Information geometry and its applications.* Tokyo: Springer. doi: 10.1007/978-4-431-55978-8

Atkinson, C., & Mitchell, A. F. (1981). Rao's distance measure. *Sankhyā: The Indian Journal of Statistics, Series A*, 345–365.

Ballou, D. (2009). Test scaling and value-added measurement. *Education Finance and Policy*, *4*(4), 351–383. doi: 10.1162/edfp.2009.4.4.351

Boothby, W. M. (1986). *An introduction to differentiable manifolds and Riemannian geometry; 2nd ed.* Orlando, FL: Academic Press.

Briggs, D. C. (2021). *Historical and conceptual foundations of measurement in the human sciences: Credos and controversies.* Routledge. doi: 10.1201/9780429275326

Calin, O., & Udrişte, C. (2014). *Geometric modeling in probability and statistics.* Springer. doi: 10.1007/978-3-319-07779-6

Caticha, A. (2022). *Entropic physics: Probability, entropy, and the foundations of physics (verion of July 26, 2022).* Retrieved from `https://www.arielcaticha.com/my-book-entropic-physics`

Feuerstahler, L. M. (2019). Metric transformations and the filtered monotonic polynomial item response model. *Psychometrika*, *84*(1), 105–128. doi: 10.1007/s11336-018-9642-9

Feuerstahler, L. M. (2020). Metric stability in item response models. *Multivariate Behavioral Research*, *57*(1), 94–113. doi: 10.1080/00273171.2020.1809980

Feuerstahler, L. M. (2023). Applications and extensions of metric stability analysis. *Psych*, *5*(2), 376–385. doi: 10.3390/psych5020025

Gelman, A., & Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, *13*(2), 163–185. doi: 10.1214/ss/1028905934

Genest, C., & Nešlehová, J. G. (2014). A conversation with James O. Ramsay. *International Statistical Review*, *82*(2), 161–183. doi: 10.1111/insr.12053

Hougaard, P. (1982). Parametrizations of non-linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *44*(2), 244–252. doi: 10.1111/j.2517-6161.1982.tb01205.x

Kass, R. E. (1989). The geometry of asymptotic inference. *Statistical Science*, 188–219. doi: 10.1214/ss/1177012480

Kass, R. E., & Vos, P. W. (2011). *Geometrical foundations of asymptotic inference*. John Wiley & Sons. doi: 10.1002/9781118165980

Lee, J. M. (2012). *Introduction to smooth manifolds* (Second ed.). Springer. doi: 10.1007/978-1-4419-9982-5

Lord, F. M. (1975). The "ability" scale in item characteristic curve theory. *Psychometrika*, *40*(2), 205–217. doi: 10.1007/BF02291567

Lord, F. M. (1980). *Applications of Item Response Theory to practical testing problems* (1st ed.). Routledge. Retrieved from `https://doi.org/10.4324/9780203056615` doi: 10.4324/9780203056615

Ly, A., Marsman, M., Verhagen, J., Grasman, R. P., & Wagenmakers, E.-J. (2017). A tutorial on fisher information. *Journal of Mathematical Psychology*, *80*, 40–55. doi: 10.1016/j.jmp.2017.05.006

McCullagh, P. (2002). What is a statistical model? *The Annals of Statistics*, *30*(5), 1225–1310. doi: 10.1214/aos/1035844977

Miyamoto, H. K., Meneghetti, F. C. C., Pinele, J., & Costa, S. I. R. (2024). On closed-form expressions for the Fisher-Rao distance. *Information Geometry*, *7*, 311–354. doi: 10.48550/arXiv.2304.14885

Myung, I. J., Balasubramanian, V., & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences*, *97*(21), 11170-11175.

Ramsay, J. O. (1996). A geometrical approach to item response theory. *Behaviormetrika*, *23*(1), 3–16.

Ramsay, J. O., & Wiberg, M. (2017). A strategy for replacing sum scoring. *Journal of Educational and Behavioral Statistics*, *42*(3), 282–302.

Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of Calcutta Mathematical Society*, *37*, 81-91.

Rossi, N., Wang, X., & Ramsay, J. O. (2002). Nonparametric item response function estimates with the em algorithm. *Journal of Educational and Behavioral Statistics*, *27*(3), 291–317.

Segert, S., & Davis-Stober, C. P. (2019). A general approach to prior transformation. *Journal of Mathematical Psychology*, *91*, 103–118.

Sijtsma, K., & Meijer, R. R. (2006). Nonparametric item response theory and special topics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of psychometrics* (pp. 719–746). Elsevier.

Tu, L. W. (2010). *An introduction to manifolds* (Second ed.). Springer.