DEBATE

https://doi.org/10.18222/eae.v36.12362

O ÍNDICE THEIL-*T* E A DIVERGÊNCIA KL: UMA NOTA COMPLEMENTAR AO PROF. R. HOFFMANN

VICTOR MAIA SENNA DELGADO[I]

[I] Universidade Federal de Ouro Preto (UFOP), Mariana-MG, Brasil; *victor.delgado@ufop.edu.br*

RESUMO

A nota responde às críticas do professor Rodolfo Hoffmann sobre o uso da divergência Kullback-Leibler (KL) para medir desigualdades educacionais. Mostra-se que o método de densidades relativas é mais geral que a medida de divergência em questão. Os exemplos de Hoffmann são questionados por usarem de comparações de distribuições uniformes e por não considerar o método de construção de Teoria da Resposta ao Item (TRI). Argumenta-se que uma boa referência para a escala de proficiência representa uma meta educacional desejada. Prova-se matematicamente, no Apêndice, que o índice de Theil-*T* é, na verdade, um caso especial da divergência KL. Isso demonstra que a KL não é uma medida inadequada, mas uma ferramenta mais geral e flexível para analisar desigualdades educacionais.

PALAVRAS-CHAVE INDICADORES EDUCACIONAIS • DESIGUALDADES EDUCACIONAIS • AVALIAÇÃO DA EDUCAÇÃO • AVALIAÇÃO DA APRENDIZAGEM.

COMO CITAR:

Delgado, V. M. S. (2025). O índice Theil-*T* e a divergência KL: Uma nota complementar ao Prof. R. Hoffmann. *Estudos em Avaliação Educacional*, *36*, Debate e12362. https://doi.org/10.18222/eae.v36.12362

EL ÍNDICE THEIL-*T* Y LA DIVERGENCIA KL: UNA NOTA COMPLEMENTARIA AL PROF. R. HOFFMANN

RESUMEN

La nota responde a la crítica del profesor Rodolfo Hoffmann sobre el uso de la divergencia Kullback-Leibler (KL) para medir las desigualdades educativas. Se muestra que el método de densidades relativas es más general que la medida de divergencia en cuestión. Los ejemplos de Hoffmann son cuestionados porque utilizan distribuciones uniformes para la comparación y por no considerar el método de construcción de la Teoría de la Respuesta al Ítem (TRI). Se argumenta que una buena referencia para la escala de aprendizaje representa un objetivo educativo deseado. Está demostrado matemáticamente, en el Apéndice, que el índice Theil-*T* es en realidad un caso especial de la divergencia KL. Esto demuestra que el KL no es una medida inadecuada, sino una herramienta más general y flexible para analizar las desigualdades educativas.

PALABRAS CLAVE INDICADORES EDUCATIVOS • DESIGUALDADES EDUCATIVAS • EVALUACIÓN DE LA EDUCACIÓN • EVALUACIÓN DEL APRENDIZAJE.

THE THEIL-*T* INDEX AND THE KL DIVERGENCE: A SUPPLEMENTARY REPLY TO PROF. R. HOFFMANN

ABSTRACT

The note responds to Professor R. Hoffmann's criticism of the use of the Kullback-Leibler (KL) divergence to measure educational inequalities. It is shown that the method of relative densities is more general than the measure of divergence in question. Hoffmann's examples are questioned because they use uniform distributions for comparison and for not considering the Item Response Theory (IRT) construction method. It is argued that a good reference for the proficiency scale represents a desired educational goal. It is proven mathematically, in the Appendix, that the Theil-*T* index is actually a special case of the KL divergence. This demonstrates that KL is not an inadequate measure, but rather a more general and flexible tool for analyzing educational inequalities.

KEYWORDS EDUCATIONAL INDICATORS • EDUCATIONAL INEQUALITIES • EDUCATIONAL EVALUATION • LEARNING ASSESSMENT.

INTRODUÇÃO

Recentemente, o distinto professor Rodolfo Hoffmann publicou neste periódico o artigo “Sobre como medir diferenças de resultados no ensino” (Hoffmann, 2025). O teor do comunicado é crítico a uma sequência de trabalhos liderados pelo professor José Francisco (Chico) Soares, da Universidade Federal de Minas Gerais (UFMG), que há anos vem trabalhando no tema de medição das desigualdades de aprendizados de alunos do ensino básico brasileiro. Em resumo, o professor Hoffmann afirma que as medidas utilizadas não podem ser aplicadas, oferece contraexemplos e enumera algumas medidas que acredita serem melhores para o propósito.

Eu participei de um dos artigos mencionados na crítica, Soares e Delgado (2016), bem como de diversos outros trabalhos direta ou indiretamente relacionados. Não participei, entretanto, do artigo analisado em maior detalhe por Hoffmann: “Desigualdades educacionais no Brasil contemporâneo: Definição, medida e resultados”, de Ernica et al. (2025).[1] Apesar de não ter participado como autor desse texto principal, creio que também me cabe o direito de resposta, dado que estive (e ainda estou) bastante envolvido com os desdobramentos da medida Kullback-Leibler (KL) adotada e dado que a “Medida das desigualdades de aprendizado entre estudantes de ensino fundamental”, de Soares e Delgado (2016), foi artigo também mencionado na crítica.

Insta ressaltar que o teor desta nota é inteiramente advindo da minha interpretação e leitura do interessante artigo crítico de Hoffmann, isentando, portanto, os demais autores citados e envolvidos em outras partes das pesquisas. Sem dúvida, também são inteiramente meus os erros e omissões em que porventura eu possa ainda estar incorrendo e que ainda remanesçam. Assumo essa responsabilidade. Não obstante não me furtarei também a apontar partes em que não concordo com os meus estimados colegas autores. Contudo penso que minhas divergências com estes são menores diante da defesa do propósito maior da pesquisa.

Considero esta nota “complementar”, pois a resposta mais importante já foi encaminhada pelos autores Soares e Ernica (2025). Resposta que conceitualmente julgo bastante pertinente. Aproveitarei, então, neste texto, para ter a liberdade de entrar em assuntos mais técnicos apontados por Hoffmann, mas que são igualmente importantes para elucidar mais aspectos da literatura abordada. Dividi esta nota em aspectos de concordância com a crítica, pontos de dissenso e as conclusões, apontando um resultado matemático que penso estar correto mostrando que a divergência KL, quando se utiliza de uma referência estritamente igualitária na dimensão da fração da renda, é exatamente o índice T de Theil – essa formulação matemática

1 Versões iniciais desse artigo já estavam disponíveis desde 2019. Tomarei como referência principalmente esse texto de Ernica et al. (2025), o mais recente. Hoffmann (2025) utiliza uma versão *preprint* de 2023; corrigi as páginas para a versão mais recente quando necessário.

está no Apêndice. Questões sobre a teoria da informação e de entendimentos do uso desses índices e indicadores na educação são apontadas ao longo do texto.

ASPECTOS DE CONCORDÂNCIA COM A CRÍTICA

É possível fazer a distinção entre argumentos *teóricos* e *práticos* das "medidas usuais de desigualdade", tais como Gini (G) e Theil (T). Ernica et al. (2025, pp. 19-20) mencionam três pressupostos que tais índices assumem. Refraseando o argumento desses autores de maneira sucinta:

i. a distribuição de referência é a igualdade de todos os rendimentos (isto é, a redundância dos n habitantes com rendimentos iguais à média, μ);[2]

ii. o somatório dos rendimentos está fixo: ($\Sigma_i^n x_i = n$);[3]

iii. o princípio de que "pode haver transferência de quantidades de renda de um indivíduo para outro, diminuindo a renda dos que concentram mais para aumentar a dos que concentram menos, de modo a se produzir situações menos desiguais".[4]

Prof. Hoffmann (2025, p. 4) argumenta que nenhum desses três pressupostos invalidaria o uso dos "consagrados" índices de Gini, Theil e outras medidas relacionadas, tanto no caso da renda quanto no caso das diferenças de resultado do ensino, o que faz o autor se contrapor à afirmação de Ernica et al. (2025, p. 20) de que "esses pressupostos, contudo, não podem ser assumidos para a realidade educacional". Primeiro, Hoffmann mostra que os estudos de desigualdade se concentram na evolução dos índices G e T, e não no referencial de G, $T = 0$, sendo que isso não deve atuar como crítica contundente às medidas de desigualdade.

Segundo, sobre o somatório de rendimentos fixos, o professor argumenta que isso não se aplica aos estudos empíricos de desigualdade de renda, pois "é óbvio que isso não ocorre quando comparamos a desigualdade de distribuição de renda no Brasil em 2001 e 2015" (Hoffmann, 2025, p. 4).

No terceiro argumento, o professor afirma que a maioria das mudanças de distribuições de renda captadas por G e T (seja para aumentos ou diminuições) não ocorre por transferências monetárias, mas sim pela dinâmica de crescimento da

2 Isso pode ser traduzido para outras unidades amostrais, tais como uma família ou domicílio. Sobre isso, ver Atkinson (2015).

3 O x_i é o valor do rendimento para cada indivíduo (ou unidade domiciliar) i, n é o tamanho da população e μ, a média. Não se propôs usar o índice de Theil para um x_i sendo a proficiência na educação por motivos que serão explicitados na próxima seção.

4 Trata-se de uma versão de transferências progressivas e mais geral do princípio de Pigou-Dalton. Hoffmann et al. (2019, p. 52) enunciam esse princípio da seguinte maneira: "A condição de Pigou-Dalton . . . estabelece que as medidas de desigualdade devem ter seus valores aumentados quando são feitas transferências regressivas de renda".

renda entre os diferentes estratos da população. Isto é, se a renda do grupo dos indivíduos mais pobres crescer mais rápido do que a renda dos indivíduos mais ricos, *G* diminuirá. Hoffmann pensa que essa dinâmica dos grupos pode ser transferida para a educação, sendo que a necessidade física da transferência (que é mesmo impossível para o caso educacional) não invalida o uso. Entende-se que os grupos que estão mais atrasados na aquisição de aprendizado (medido por testes padronizados) podem fazer o *catching up*, e assim a desigualdade de aprendizado, medida tanto por grupos quanto individualmente, diminuiria.

Os três contrapontos levantados por Hoffmann (2025) são relevantes, pois são calcados na prática consolidada dos indicadores de desigualdades aplicados à renda, que em muitos casos podem ser transpostos para a análise da desigualdade de aprendizado. Tive a oportunidade de ler a prévia da resposta de Soares e Ernica (2025), na qual os autores defendem que os pressupostos envolvidos têm implicações diretas nos direitos fundamentais dos estudantes. Em outras palavras, uma boa referência é mais adequada quando aplicada ao aprendizado educacional, já que é nesse âmbito que se pode converter em políticas públicas efetivas. Concordo com essa interpretação e tratarei dela novamente na próxima seção.

Além desses pontos, em que se pode ter uma diferença na veemência sobre o quanto eles influenciam na mais justa interpretação (ou não) do problema, há outros dois pontos menores, embora interessantes, apontados pelo autor (e com os quais concordo):

1. Na nota de rodapé n. 4, Hoffmann escreve o seguinte:

   O fato de a divergência de Kullback-Leibler nunca ser negativa é intrínseca ao seu significado dentro da teoria da informação. O valor informativo da previsão que transforma a distribuição Q na distribuição P nunca é negativo. *Não nos parece válido manter o mesmo nome para a medida depois de lhe adicionar um sinal que pode ser negativo*. Além disso, Soares et al. (2018, p. 14) dizem que o sinal vai depender da posição da curva de distribuição acumulada relativa em relação à linha de 45°, sem mencionar o que fazer se a curva e a linha se cruzarem. (Hoffmann, 2025, p. 10, grifo nosso).

2. Na página 11, é mencionado que, "considerando que *escolaridade* e notas são *variáveis ordinais*, deveríamos usar a mediana, e não a média, como medida de tendência central da distribuição" (Hoffmann, 2025, p. 11, grifo nosso).

Não poderia concordar mais com essa nota n. 4 de Hoffmann. Como autor do referido trabalho, apenas me penitencio por não conseguir apresentar aos meus estimados colegas coautores nenhuma outra solução que os convencesse a não modificar a medida KL. Sob o risco de não cairmos em um *ad hoc*, creio que acabamos caindo em outro. Novas interpretações ou soluções são bem-vindas; hoje considero que tenho algumas e que vão na linha do conjunto de medidas que Hoffmann propõe após sua crítica – expô-las aqui, no entanto, ocuparia um espaço que não é o propósito deste texto.

Sobre escolaridades e notas serem variáveis *ordinais*, bem como a interessante discussão sobre escala razão, penso que, como diz o ditado, o professor Hoffmann "atirou no que viu e acertou no que não viu". Em momento algum nosso crítico mencionou que a *aprendizagem* medida pelos escores de proficiência advém do método da Teoria da Resposta ao Item (TRI). O assunto da cardinalidade ou ordinalidade das unidades de medida tem raízes profundas na teoria econômica (e também em estatísticas sociais); não convém adentrar aqui. Sobre a TRI é importante mencionar Hambleton et al. (1991) e os pesquisadores nacionais Klein (2013), Pasquali e Primi (2003) e Fletcher (1994). Entretanto o método de construção da escala de proficiência é crucial para entender a proposta de Soares e Delgado (2016) e de Ernica et al. (2025). Sem entender isso, todo o esforço da distribuição de referência se torna vazio.

As proficiências derivadas de variáveis latentes geram, por construção, uma distribuição normal dos alunos, atribuindo a cada estudante um escore que pode posicioná-lo de forma mais ou menos favorável, em função da *randomização* na seleção dos itens da prova. Embora exista certo grau de subjetividade no método de construção, os exemplos apresentados por Hoffmann (seções 2-4) sugerem uma adesão persistente à Teoria Clássica dos Testes (TCT) e à lógica do julgamento subjetivo de notas, o que reforça a relevância da ordinalidade em propósitos classificatórios. Porém reforça-se aqui que a TCT tem limitações já há muito superadas pela TRI.

Por fim, ressalto que, ainda que em tom crítico, é digno de reconhecimento o esforço do professor Hoffmann em analisar de forma abrangente os mais de cinco artigos relacionados ao tema, sintetizando-os com clareza didática e incorporando elementos de sua experiência e *expertise*. Para além dos pontos destacados nesta seção, sua exposição revela domínio notável da teoria da informação, da divergência de Kullback-Leibler, das implicações das medidas propostas e da complexa estatística envolvida. Trata-se, portanto, de uma contribuição que reflete a profundidade de sua trajetória acadêmica e intelectual.

PONTOS DE DISSENSO

Há que se contrapor a contundência da crítica. Como se pode intuir da seção anterior, há uma questão de interpretação que passou ao largo da avaliação de nosso crítico. Não é preciso recorrer ao "tudo vale a pena . . .", do poeta Fernando Pessoa, para mostrar que a frase "não vemos nenhuma vantagem em utilizar a divergência de Kullback-Leibler na análise das notas" (Hoffmann, 2025, p. 9) ou o resumo que abre a conclusão – "Nossa recomendação é abandonar totalmente o uso da divergência de Kullback-Leibler na análise das notas" (p. 18) – são recomendações demasiadamente intransigentes.

Neste volume, a resposta de Soares e Ernica (2025) trata de forma enfática a parte da interpretação da referência e de suas consequências no direito à *igualdade* dos grupos de estudantes. Sendo assim, me aterei a alguns aspectos mais técnicos e que inclusive servem de título a esta nota.

Os diversos contraexemplos fornecidos por Hoffmann (2025) não são do tipo que refutam a nossa argumentação original. Por dois critérios a serem elucidados: a) se a referência utilizada é uma distribuição de densidade uniforme, isso indica que não importa onde a discrepância ocorre (trabalharemos o exemplo rapidamente adiante); b) é importante trabalhar com algumas questões pertinentes à medição da proficiência, tal como o método TRI, mencionado na seção anterior. A referência é importante.

A divergência de Kullback e Leibler (1951) surgiu no contexto da teoria da informação de Shannon. Seguindo a notação de nosso crítico, para duas distribuições de probabilidades, sempre positivas, chamadas de $P(x_i)$ e $Q(x_i)$ para uma variável x_i, que possui grupos $i = \{1, \ldots, n\}$, a divergência KL para o caso discreto se torna:

$$KL(P \parallel Q) = \sum_{i=1}^{n} P(x_i) \ln \left(\frac{P(x_i)}{Q(x_i)} \right) \tag{1}$$

Essa medida resume, portanto, o quanto uma P é discrepante de Q. Se, para todos os valores de i, P e Q são iguais, teremos $KL = 0$, que resume que não há discrepância entre os dois perfis de probabilidade.

Para contrapor os contraexemplos de Hoffmann, vale ressaltar que a medida não distingue de qual x_i advém a divergência, no sentido de que não importa se os *bits* da informação discrepante (*bits* no caso de uma KL com $\log_2$) advêm do primeiro, segundo ou n-ésimo indivíduo i.

Sendo assim, os quatro exemplos de Hoffmann (2025, pp. 8-9) não são muito úteis, realmente, pois as três primeiras distribuições, aqui denominadas P_1, Q_1 e U_1, são comparadas com duas medidas, $KL(Q_1 \parallel P_1)$ e $KL(U_1 \parallel P_1)$, nas quais P_1 é perfeitamente uniforme. Ora, ser uniforme indica que não importa de onde vem a discrepância. Mais adiante, ainda na mesma seção, Hoffmann (pp. 9-10) apresentou as distribuições aqui batizadas de P_2, Q_2 e U_2, que possuem maior variância, mas muitos grupos com a mesma probabilidade para a obtenção de diferentes notas; isso influencia a comparação das distribuições (ver Tabela 1, adiante).

Adaptei os seis perfis de distribuição de Hoffmann (2025) para comparar com uma função de referência R que capta o espírito de Soares e Delgado (2016) e de Ernica et al. (2025). As três últimas linhas da Tabela 1 apresentam, respectivamente, a divergência KL, o índice de Gini e o índice *T* de Theil. Os zeros foram tratados da seguinte maneira: 0·log(0) = 0, o que subestima um pouco as desigualdades, mas torna *T* empiricamente calculável para casos com $x_i = 0$. Nota-se que as probabilidades das notas de *R* são crescentes de 0 a 10. Isso denota que a referência é uma construção *normativa*.

TABELA 1

Distribuições de frequências fictícias sugeridas por Hoffmann e uma referência *R* proposta nesta nota

NOTA x_i	REFERÊNCIA (*R*)	EXEMPLOS DA TABELA 1			EXEMPLOS DA TABELA 2		
		U_1	P_1	Q_1	U_2	P_2	Q_2
0	10^{-7}	0,054	0,091	0,054	0,05	0,05	**0,10**
1	10^{-6}	0,054	0,091	0,054	0,05	0,05	**0,40**
2	10^{-4}	0,054	0,091	0,054	0,05	0,05	**0,10**
3	0,001	0,054	0,091	**0,454**	0,05	0,05	0,05
4	0,005	0,054	0,091	0,054	0,05	0,05	0,05
5	0,010	0,054	0,091	0,054	0,05	**0,10**	0,05
6	0,020	0,054	0,091	0,054	0,05	**0,40**	0,05
7	0,050	0,054	0,091	0,054	0,05	**0,10**	0,05
8	0,150	0,054	0,091	0,054	**0,10**	0,05	0,05
9	0,300	0,054	0,091	0,054	**0,40**	0,05	0,05
10	0,464	**0,454**	0,091	0,054	**0,10**	0,05	0,05
KL	0,000	2,002	3,669	4,458	1,866	3,061	7,413
G	0,050	0,220	0,270	0,260	0,200	0,170	0,420
T	0,009	0,165	0,247	0,205	0,143	0,130	0,454

Fonte: Elaboração do autor, com base em Hoffmann (2025).

Notas: (1) Adaptado das distribuições *P*, *Q* e *U* de Hoffmann (2025, pp. 8-9), tabelas 1 e 2, com distribuição *R* proposta neste texto para referência e as medidas comparativas. Valores discrepantes em negrito para melhor visualização.

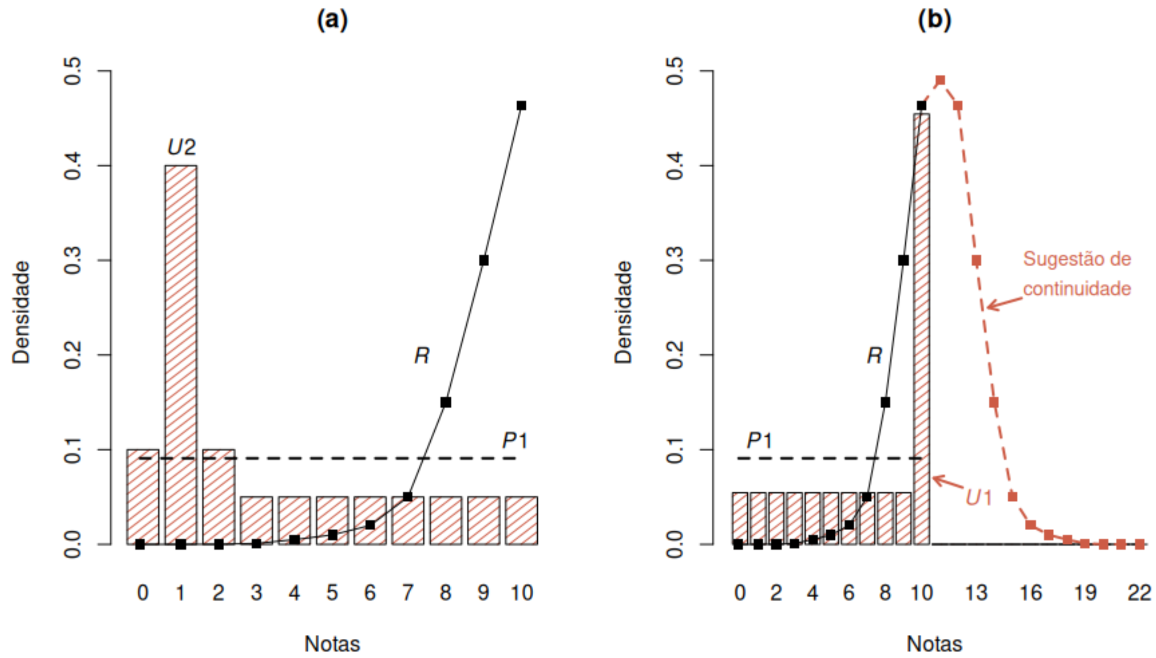
(2) "Tabela 1" e "Tabela 2", aqui, referem-se às tabelas que constam do artigo de Hoffmann (2025).

O gráfico da Figura 1 mostra a ideia para a obtenção dos valores da referência. O ponto é que *R* apresenta excelente desempenho (média 9,083 e variância 2,29), que é uma média maior e uma variância menor do que a de todos os exemplos da seção 2 de Hoffmann (2025). Ou seja, é uma referência positiva para melhoria de desempenho dos estudantes (Figura 1-a), algo pretendido. Mais ainda, na escala de proficiência TRI do Sistema de Avaliação da Educação Básica (Saeb), que não é limitada entre um valor inferior e um superior, é possível verificar valores mais altos do que os comumente observados – essa ideia é representada conceitualmente pela linha pontilhada da continuação de *R* no gráfico da Figura 1-b, na qual se apresenta *sugestão de continuidade*, que é algo que ocorre na escala Saeb transposta de Soares e

Delgado (2016); a referência é transposta, ainda, para uma escala de valores válidos da escala Saeb. Entretanto, no exemplo de Hoffmann, em que 0 é o valor inferior e 10 o superior, isso não é possível.

Deslocar as densidades para além do valor 10, logicamente, requereria ajuste nos valores sugeridos no exemplo e uma escala sem limite superior estrito (ou seja, sem máximo teórico, (**inf**, **sup**) = (-∞,+∞), apenas o empírico).

FIGURA 1
Ideia conceitual da distribuição de referência *R*



Fonte: Elaboração do autor, com base em Hoffmann (2025).
Nota: As distribuições P_1, U_1 e U_2 são os exemplos de Hoffmann (2025, pp. 8-9). A referência *R* é a construção proposta neste texto, e a "sugestão de continuidade" é para interpretação de escalas não limitadas no intervalo 0 a 10.

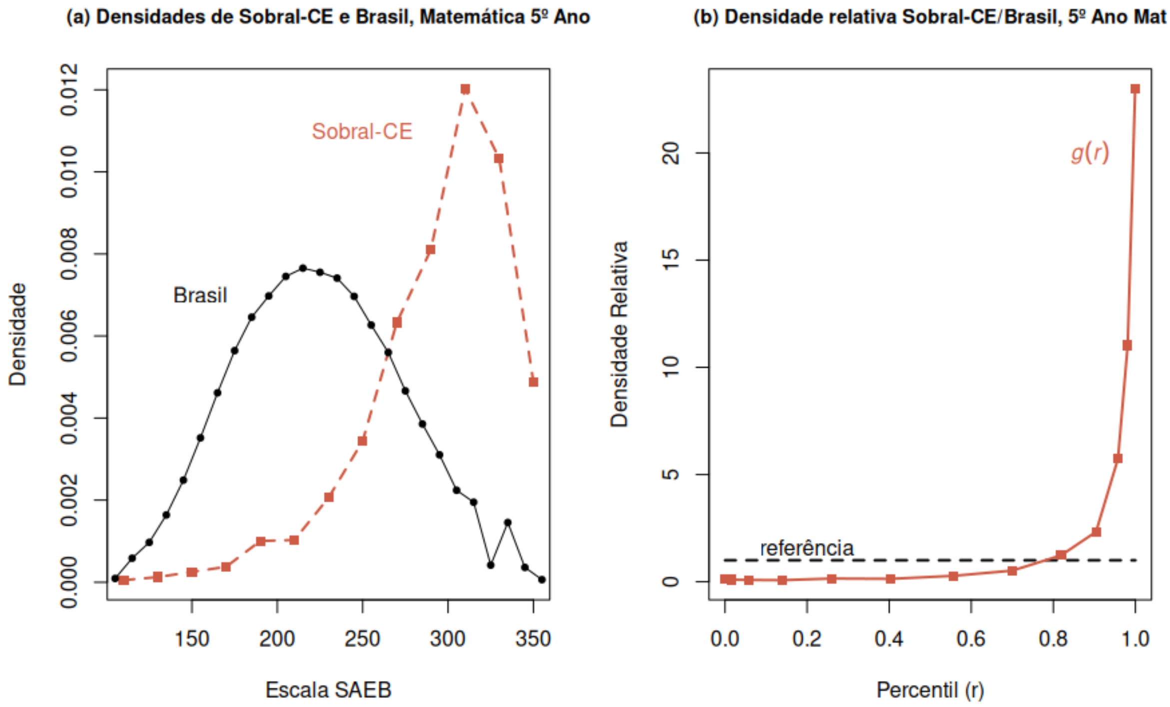
O esforço do professor J. F. Soares é para obter uma referência internacional e de qualidade para a escala Saeb do ensino brasileiro. Tal proposta surge quase naturalmente ao se notar a distância da distribuição do Brasil para outros países de referência no Programme for International Student Assessment (Pisa). Essa ideia é central no artigo de Soares e Delgado (2016).

A referência não precisa ser apenas aquela sugerida pelos autores do artigo original. Tal distribuição pode, por exemplo, se basear em casos observados, como o bem documentado caso de Sobral-CE (Cruz & Loureiro, 2020). O risco em que se incorre ao ancorar a referência em uma distribuição "empírica" é não ter a sugestão de um potencial de desempenho superior ao realizado. Preocupação que se manifesta nos trabalhos sobre os níveis do Saeb com o uso das técnicas aqui discutidas.

Na Figura 2-a, a seguir, apresenta-se a comparação do desempenho em matemática no Saeb 2019 do 5º ano entre Brasil e Sobral. Verifica-se que a distribuição de Sobral concentra maior densidade em níveis superiores, o que resulta em uma função $g(r)$ convexa na Figura 2-b. Discrepâncias são mais evidentes na cauda superior, isto é, entre os alunos de maior rendimento. A medida KL obtida (0,058) aproxima-se daquela registrada por Soares e Delgado (2016) para o 5º ano em matemática. Sugestões acerca de medidas descritivas, sumários, propriedades de referência e suas respectivas vantagens e limitações são bem-vindas.

FIGURA 2

Comparativo de desempenho em matemática (Saeb, 5º ano, 2019) entre Brasil e Sobral-CE



Fonte: Elaboração do autor, com base nos dados da Prova Brasil 2019 (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [Inep], 2020).

Note-se que tanto Gini quanto Theil são medidas integrais de dispersão, comparam apenas esse aspecto das distribuições. Já a medida KL capta diferenças tanto na localidade da distribuição quanto na sua dispersão. Handcock e Morris (1999, cap. 3) mostram como tais decomposições podem ser realizadas. Seja X a distribuição de notas observada, a distribuição R pode ser interpretada como o Data Generating Process (*DGP*) *normativo* que se pretende para o sistema educacional básico brasileiro. Com isso, $KL(X \,||\, R)$ se torna o valor de discriminação (o esforço de entropia) para sair do sistema observado X para o sistema ideal R; quanto maior esse valor, maior o esforço necessário.

O uso das divergências de Kullback-Leibler e do método mais geral de *densidades relativas* em variáveis sociais, no Brasil, iniciou-se em trabalhos do Centro de Desenvolvimento e Planejamento Regional da Universidade Federal de Minas Gerais (Cedeplar/UFMG), inspirados no artigo de Handcock e Morris (1998) e no livro subsequente (1999). Essa linha de pesquisa gerou dissertações, teses e artigos aplicados à renda e à educação.[5] Meu contato com a técnica ocorreu durante o doutorado (2009-2013), quando o professor Soares, então pesquisando referências para ancoragem do Saeb e atuando em bancas no Cedeplar, sugeriu seu uso, o que resultou no artigo Soares e Delgado (2016) e em alguns trabalhos prévios, que se iniciaram em 2010.

Breve comentário sobre as seções 3 e 4 de Hoffmann

Considero pertinente o esforço de Hoffmann (2025), especialmente nas seções 3 e 4, ao utilizar dados da Pesquisa Nacional por Amostra de Domicílios (Pnad) Contínua de 2022 (Instituto Brasileiro de Geografia e Estatística [IBGE], 2023). Tais dados foram utilizados para mensurar a escolaridade e aplicar a divergência KL a uma distribuição de referência que o professor julgou adequada ao exercício. Contudo, como o vínculo entre escolaridade e educação não constitui preocupação central deste trabalho – cujo foco recai sobre a compreensão da desigualdade educacional, tema de particular relevância para o Brasil –, limitar-me-ei a comentar essas seções de forma breve.

Hoffmann (2025, p. 13) escreve que sua Equação 5 "mostra que variações na escolaridade correspondem a variações relativas na renda, ou seja, é a dispersão na escolaridade, e não a desigualdade da escolaridade, que está diretamente associada à desigualdade na renda". Apesar de a relação minceriana (ver Heckman et al., 2006) ter sido assumida de antemão na Equação 4 (de Hoffmann), não questiono nem contraponho a frase desse parágrafo. O ponto da sequência de trabalhos defendidos é outro: o de que as desigualdades educacionais são uma questão de direito e atrasam a aquisição educacional de determinados grupos de alunos, o que deixa o Brasil sempre em atraso no que diz respeito à aquisição de aprendizado.

Outro ponto que convém ser destacado é que o método de distribuições relativas de Ali e Silvey (1966), Ćwik e Mielniczuk (1989) e Handcock e Morris (1998, 1999) é mais geral do que a medida sintética de Kullback-Leibler. Tal método vai inclusive no sentido geral da recomendação de Hoffmann, que fala sobre a dificuldade de

5 É difícil acompanhar todos os trabalhos da UFMG que utilizaram a técnica de densidade relativa, mas recordo a tese de Clarissa G. Rodrigues (2009) (orientada pelo professor Eduardo L. G. Rios-Neto), cujo *paper* Rodrigues et al. (2013) é um trabalho relacionado. Antes disso, é possível citar também a argumentação de Raquel R. de M. Guimarães (cujo trabalho de conclusão de curso em tema correlato foi orientado pela professora Ana Hermeto de Oliveira), que aparece em Oliveira e Guimarães (2009).

encontrar medidas sintéticas (ver Figura A1 no Apêndice deste texto). Ali e Silvey (1966) provam que tal medida é suficiente no sentido estatístico, isto é, carrega toda a informação necessária para a obtenção dos parâmetros descritores de P e Q.

O gráfico da distribuição relativa é elucidativo, em especial, quando as distribuições estão próximas, mostrando as possíveis fontes de divergência. Shlens (2014) mostra que a KL é resultado da esperança da função de verossimilhança de observar P quando o *DGP* verdadeiro é Q. Outra forma de ver é a interpretação de que a medida KL é a média da distribuição relativa $g(r)$ ponderada por uma das distribuições (ver Equação A6 no Apêndice).

Observa-se, novamente, que, embora a aquisição da escolarização, medida em anos de estudo, seja uma variável importante para a economia e para o desenvolvimento econômico, não é a preocupação dos trabalhos que medem a desigualdade de aprendizado por meio das proficiências. Na verdade, acredita-se, por demais evidências da área, que os brasileiros se escolarizam pouco por terem uma aquisição educacional de baixo aprendizado.

Além disso, as medidas TRI de variável latente de aprendizado (tanto a escala Saeb quanto a escala Pisa) possuem distribuição normal, e existe um esforço importante para a comparabilidade de seus valores. Wallmark et al. (2024) tentam inclusive relacionar as escalas TRI com "*escalas de razão*" indicadas por Hoffmann. É um esforço de fronteira na ciência, mas pode render frutos. Outro ponto relevante é que é possível ponderar que a variável *renda*, medida por unidades monetárias, é tampouco "*escala razão*", pois são conhecidas as limitações da subjetividade do valor. No entanto isso não impede o uso geral dos índices de Gini, Theil e congêneres. A TRI constitui, de fato, uma escala intervalar (Organisation for Economic Co-operation and Development [OECD], 2023, 2009), embora com suas dificuldades e questões inerentes.

Os últimos anos têm evidenciado como a sociedade brasileira está preocupada com a desigualdade educacional entre os grupos, conforme assinala o próprio trabalho de Ernica et al. (2025). Essa desigualdade impacta na aquisição educacional medida por anos de escolaridade, mas em primeiro momento é preciso compreendê-la nos pormenores, por conta das consequências desse conhecimento para a política pública educacional. É isso que Soares e Ernica (2025) advogam, e eu faço companhia a esses autores.

CONCLUSÕES

A recomendação final de Hoffmann (2025, p. 18) de "abandonar totalmente o uso da divergência de Kullback-Leibler na análise de notas" parece-me evidentemente exagerada. Como argumentado, o método de *densidades relativas* é mais amplo do que a medida sintética da divergência KL e está amplamente referendado pelos diversos trabalhos acadêmicos mencionados neste texto, além de que a medida KL foi perfeitamente justificada pelos autores que dela fizeram uso.

Note-se que o uso de qualquer medida sintetizadora pode, com certeza, ser questionado. Mais ainda pelo fato de o Prof. Hoffmann acreditar que não trabalhávamos com uma medida de escala razão. No entanto isso apenas denota o desconhecimento do professor quanto a renomados esforços dos estudos de TRI para ter uma medida que permita comparação entre países, aos anos de aplicação dos *surveys* e diferentes estratos sociais dos alunos. Embora exista o questionamento válido se as escalas de proficiência deixam de ser uma medida não ordinal, pois hoje já se sabe que são influenciadas por diversas variáveis de contexto e não estão isentas de problemas, os esforços de anos de Pisa e do Saeb são admiráveis e não foram contestados por inúmeros pesquisadores de renome que já se debruçaram sobre a área até o momento.

Notadamente, são necessários esforços para aprofundar o conhecimento sobre o uso da divergência KL no campo educacional, quando for pertinente. Além disso, é importante compreender de que forma o método de densidades relativas pode contribuir, bem como identificar os parâmetros descritivos que essa técnica permite obter.

Ainda, como indicam Ernica et al. (2025), a desigualdade relevante entre os alunos do ensino básico brasileiro se dá entre grupos definidos por características sociais descritivas, sendo a desigualdade advinda de condições de nascimento (sexo, raça ou região de origem e nível socioeconômico) e identitárias (gênero, religião, crenças e valores), aquelas que requerem maior explicação. É sobre essas questões que se devem debruçar as medidas de desigualdade de aprendizado da educação. Isso, no fundo, envolve conhecer por que determinada aluna ou aluno (no caso individual), ou grupo social (no caso coletivo), não consegue alcançar os patamares mínimos de aprendizado e com isso vê vetados os seus direitos.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Código de Financiamento 001, e do suporte provido pela Universidade Federal de Ouro Preto (UFOP). Agradeço a leitura crítica e o incentivo para continuar de J. F. Soares, Eduardo L. G. Rios-Neto, Matheus Gomes, Murilo Valim, Marcelo A. C. Nogueira e Fernanda Faria Silva, bem como aos pareceristas e revisores anônimos da *Estudos em Avaliação Educacional*, que ajudaram a melhorar este texto. Erros remanescentes são de minha inteira autoria.

REFERÊNCIAS

Ali, S. M., & Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological), 28*(1), 131-142. https://www.jstor.org/stable/2984279

Atkinson, A. B. (2015). *Inequality: What can be done?* Harvard University Press.

Cruz, L., & Loureiro, A. (2020). *Alcançando um nível de educação de excelência em condições socioeconômicas adversas: O caso de Sobral*. World Bank Group. http://documents.worldbank.org/curated/en/778741594193637332

Ćwik, J., & Mielniczuk, J. (1989). Estimating density ratio with application to discriminant analysis. *Communications in Statistics – Theory and Methods, 18*(8), 3057-3069. https://doi.org/10.1080/03610928908830077

Ernica, M., Rodrigues, E. C., & Soares, J. F. (2025). Desigualdades educacionais no Brasil contemporâneo: Definição, medida e resultados. *Dados, 68*(1), Artigo e20220109. https://doi.org/10.1590/dados.2025.68.1.345

Fletcher, P. R. (1994). A Teoria da Resposta ao Item: Medidas invariantes do desempenho escolar. *Ensaio: Avaliação e Políticas Públicas em Educação, 1*(2), 21-27. http://educa.fcc.org.br/pdf/ensaio/v01n02/v01n02a04.pdf

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage.

Handcock, M. S., & Morris, M. (1998). Relative distribution methods. *Sociological Methodology, 28*(1), 53-97. https://doi.org/10.1111/0081-1750.00042

Handcock, M. S., & Morris, M. (1999). *Relative distribution methods in the social sciences*. Springer.

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In E. Hanushek, & F. Welch (Eds.), *Handbook of the economics of education* (Vol. I; pp. 307-458). North Holland. https://doi.org/10.1016/S1574-0692(06)01007-5

Hoffmann, R. (2025). Sobre como medir diferenças de resultados no ensino. *Estudos em Avaliação Educacional, 36*, Artigo e10663. https://doi.org/10.18222/eae.v36.10663

Hoffmann, R., Botassio, D. C., & Jesus, J. G. de. (2019). *Distribuição de renda: Medidas de desigualdade, pobreza, concentração, segregação e polarização*. Edusp.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2023). *Pnad Contínua – Pesquisa Nacional por Amostra de Domicílios Contínua 2022: Microdados*. IBGE. https://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Trimestral/Microdados/2022/

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2020). *Microdados do Saeb 2019*. Inep. https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/saeb

Klein, R. (2013). Alguns aspectos da Teoria de Resposta ao Item relativos à estimação das proficiências. *Ensaio: Avaliação e Políticas Públicas em Educação, 21*(78), 35-56. https://doi.org/10.1590/S0104-40362013005000003

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79-86. https://www.jstor.org/stable/2236703

Oliveira, A. M. H. C. de, & Guimarães, R. R. de M. (2009). Trends in the relative distribution of wages by gender and cohorts in Brazil (1981-2005). In *Proceedings of the Population Association of America 2009 Annual Meeting* (pp. 1-21). https://paa2009.populationassociation.org/papers/90744

Organisation for Economic Co-operation and Development (OECD). (2009). *PISA Data Analysis Manual: SPSS®* (2nd ed.). Pisa; OECD Publishing. https://doi.org/10.1787/9789264056275-en

Organisation for Economic Co-operation and Development (OECD). (2023). *PISA 2022 results: Learning during – and from – disruption* (Vol. II). Pisa; OECD Publishing. https://doi.org/10.1787/a97db61c-en

Pasquali, L., & Primi, R. (2003). Fundamentos da Teoria da Resposta ao Item – TRI. *Avaliação Psicológica, 2*(2), 99-110. https://pepsic.bvsalud.org/pdf/avp/v2n2/v2n2a02.pdf

Rodrigues, C. G. (2009). *A relação entre a expansão do acesso ao ensino e o desempenho escolar no Brasil: Evidências com base no Saeb para o período de 1997 a 2005* [Tese de doutorado, Universidade Federal de Minas Gerais]. Repositório Institucional UFMG. https://hdl.handle.net/1843/AMSA-84CKF5

Rodrigues, C. G., Rios-Neto, E. L. G., & Pinto, C. C. de X. (2013). Changes in test scores distribution for students of the fourth grade in Brazil: A relative distribution analysis for the years 1997-2005. *Economics of Education Review, 34,* 227-242. https://doi.org/10.1016/j.econedurev.2012.12.006

Shlens, J. (2014). Notes on Kullback-Leibler divergence and likelihood. *ArXiv*. https://doi.org/10.48550/arXiv.1404.2000

Soares, J. F., & Delgado, V. M. S. (2016). Medida das desigualdades de aprendizado entre estudantes de ensino fundamental. *Estudos em Avaliação Educacional, 27*(66), 754-780. https://doi.org/10.18222/eae.v27i66.4101

Soares, J. F., & Ernica, M. (2025). O conceito rege o método: Uma resposta a Hoffmann. *Estudos em Avaliação Educacional, 36,* Debate e12270. https://doi.org/10.18222/eae.v36.12270

Wallmark, J., Josefsson, M., & Wiberg, M. (2024). Introducing flexible monotone multiple choice Item Response Theory models and bit scales. *ArXiv*. https://doi.org/10.48550/arXiv.2410.01480

APÊNDICE

Apêndice A – Relação entre a divergência KL e o índice de Theil

Hoffmann et al. (2019, p. 98) relacionam o índice T de Theil com a medida de informação de uma mensagem incerta:

$$I(x \parallel y) = \sum_{i=1}^{n} y_i \ln \frac{y_i}{x_i} \tag{A1}$$

em que y_i é a fração da renda apropriada para cada grupo i, chamada de *a posteriori*, e x_i é a informação *a priori*. O n é o número de indivíduos da sociedade observada, advindo de um conjunto de indivíduos $\{1,2, \dots ,n\}$. No caso da medida T, *a priori* é a fração da renda a ser igualmente dividida entre os n indivíduos, $1/n$. O ganho de informação (ou discrepância da informação) ocorre na divergência da observação das frações de fato apropriadas pelos indivíduos (y_i) e daquelas que seriam adequadas *normativamente* pela distribuição igualitária ($x_i = 1/n$).

Nota-se a semelhança da Equação A1 com a Equação 1 deste texto. Segue a prova por construção de que Theil é igual à divergência Kullback-Leibler para o caso específico em que a probabilidade $P(x_i)$ é dada pela fração apropriada da renda e $Q(x_i)$ a apropriação de renda na sociedade igualitária.

Tomemos y_i como proveniente de uma variável aleatória Y, que possui a probabilidade de ocorrência igual à fração de renda apropriada de cada indivíduo i (ou grupo i). Isto é, y_i é uma realização observada de Y. Comparemos essa variável com outra variável aleatória X, que, no caso, tem probabilidade equânime e igual a $1/n$, isto é, uma uniforme em que cada realização de x_i é igual a $(1/n)$.

Por fim, defina-se $\boldsymbol{I}^2$ como o espaço de quadrado unitário em que ambas as variáveis, X e Y, estejam definidas. Sendo que X: $[0,1] \rightarrow 1/n$ e Y advém de uma função mais geral em que f: $[0,1] \rightarrow [0,1]$. Então, podemos definir a seguinte proposição.

Proposição 1. Dados $\boldsymbol{I}^2$ e n, o índice T de Theil é a divergência KL para o caso específico da função de fração da renda y_i e o referencial $x_i = 1/n$.

Demonstração. Seja o índice de Theil-T dado por:

$$T = \sum_{i=1}^{n} \left[y_i \ln \left(\frac{y_i}{x_i} \right) \right] \tag{A2}$$

Definindo-se x_i advindo da função constante X: $[0,1] \rightarrow 1/n$, obtém-se novamente o T de Theil:

$$T = \sum_{i=1}^{n} \left[y_i \ln \left(\frac{y_i}{(1/n)} \right) \right] \tag{A3}$$

A divergência KL com as mesmas duas variáveis Y e X com as descrições dadas acima e do quadrado unitário $\boldsymbol{I}^2$ possui a seguinte formulação:

$$KL(Y \parallel X) = \sum_{i=1}^{n} \left[y_i \ln \left(\frac{y_i}{(1/n)} \right) \right] \tag{A4}$$

Como o lado direito das duas definições A3 e A4 é igual, temos que

$$T = KL(Y \,||\, X) \qquad \text{(A5),}$$

para $X, Y \in \boldsymbol{I}^2$, com y_i sendo a fração da renda e $x_i = 1/n$ para todo $i \in \{1, \dots, n\}$. Por construção. *CQD*.

As condições de probabilidades para X e Y estão atendidas:

i. x_i, $y_i > 0$;

ii. $\Sigma x_i = \Sigma y_i = 1$.

Conclui-se que a KL é um caso mais geral, dado que o domínio típico da divergência não é o da fração da renda total apropriada pelas pessoas (a variável *Y*) comparada com o caso estritamente igualitário (variável *X*), mas, sim, tipicamente, uma densidade probabilidade, ou mesmo funções gerais *f* e *g*. Pode, portanto, ser comparada com outras densidades probabilidade; essa é a chave correta de interpretação da referência para educação.

Abaixo, a Tabela A1 reconsidera as três últimas linhas da Tabela 1 deste texto e calcula uma KL que compara as distribuições sugeridas (*R*,*P*, *Q* e *U*) com 1/*n* em cada um dos casos. O resultado é o mesmo de *T* apresentado pela Tabela 1 deste texto.

É possível transformar a divergência KL discreta da Equação 1 com a sugestão da probabilidade relativa de Handcock e Morris (1999). Suponha $g_i = P(x_i)/Q(x_i)$, que advém da multiplicação de $Q(x_i)/Q(x_i)$ na Equação 1 deste trabalho. Sendo assim, a KL se torna:

$$KL(P \parallel Q) = \sum_{i=1}^{n} g_i \ln{(g_i)} q_i = E_Q[g_i \ln{(g_i)}] \tag{A6}$$

TABELA A1
Comparação com a KL com referência igualitária 1/*n*

NOTA x_i	REFERÊNCIA (*R*)	EXEMPLOS DA TABELA 1			EXEMPLOS DA TABELA 2		
		U_1	P_1	Q_1	U_2	P_2	Q_2
KL	0,000	2,002	3,669	4,458	1,866	3,061	7,413
G	0,050	0,220	0,270	0,260	0,200	0,170	0,420
T	0,009	0,165	0,247	0,205	0,143	0,130	0,454
KL(*Y* || *X*)	**0,009**	**0,165**	**0,247**	**0,205**	**0,143**	**0,130**	**0,454**

Fonte: Elaboração do autor, com base em Hoffmann (2025).

Notas: (1) Adaptado das distribuições *P*, *Q* e *U* de Hoffmann (2025, pp. 8-9), tabelas 1 e 2, com distribuição *R* proposta neste texto para referência e as medidas comparativas. Note que *KL*(*Y* || *X*) produz os mesmos resultados de *T*. O *Y* deve ser substituído pela distribuição da coluna em medição e *X* = 1/*n*, sempre.

(2) "Tabela 1" e "Tabela 2", aqui, referem-se às tabelas que constam do artigo de Hoffmann (2025).

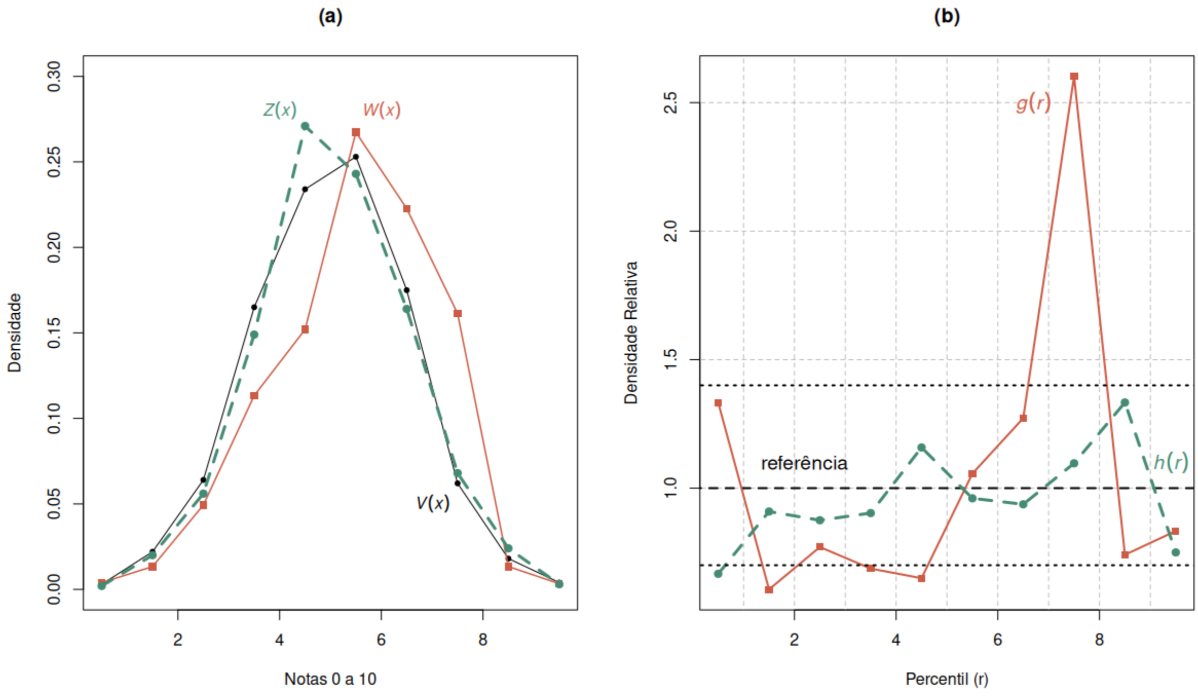
Em que q_i abrevia $Q(x_i)$. Temos então o somatório da probabilidade relativa ponderada às probabilidades de comparação *Q*; a divergência de *P* para *Q* é a surpresa esperada, em que a esperança é tomada sobre a distribuição de referência *Q*.

Para exemplificar seu uso prático, suponha três distribuições de frequências *V*(*x*), *W*(*x*) e *Z*(*x*), todas dadas por variáveis aleatórias *X* com distribuição normal, média 5 e desvio padrão 1,5, porém *W*(*x*) possui ⅓ das notas dadas por uma aleatória uniforme com valores mínimo e máximo de 5 e 8. A Figura A1 exemplifica o uso das densidades relativas.

Na Figura A1-a, as densidades são comparadas (preferiu-se aqui não "suavizar" nenhuma das distribuições para facilidade de obtenção da distribuição discreta). O perfil das densidades comparando W e V denota um acúmulo maior de densidades entre 5 e 8 em W. Porém é na Figura A1-b, na $g(r) = W(x)/V(x)$, que a diferença é notada de maneira mais acintosa. O mesmo não ocorre com $h(r) = Z(x)/V(x)$, dado que tanto V quanto Z foram obtidas pelo mesmo procedimento. A divergência $KL(W \,||\, V) = 0{,}092$ e $KL(Z \,||\, V) = 0{,}006$.

FIGURA A1

Exemplo, três distribuições normais, *V*, *W* e *Z*, com média 5 e desvio padrão 1,5. *W* é levemente alterada com ⅓ da distribuição proveniente de uma uniforme *U*[5,8]



Fonte: Elaboração do autor.

Desse modo, mostra-se como a técnica de densidades relativas é sensível o suficiente para captar que um terço do *profile W* vem de um *DGP* diferente.