arXiv:2601.00773v1 [stat.ME] 2 Jan 2026

广义线性模型中的变量重要性——使用 Shapley 值的统一视角

SinanAcemoglu* Christian Kleiber † Jörg Urban‡

2026年1月5日

摘要

回归分析中的变量重要性在许多领域都备受关注。目前尚无评估变量重要性的唯一方法。不过，现有文献中相当一部分研究会显式或隐式地使用 Shapley 值，对适当的拟合优度度量进行分解；在线性回归模型中，通常采用经典的 R^2。在线性回归之外，尚无公认的拟合优度度量，只有各种伪-R^2。我们阐述并讨论拟合优度度量应具备的理想属性，以便将 Shapley 值解释为相对重要性，乃至绝对重要性。我们建议使用一种基于库尔贝克-莱布勒散度的伪-R^2，即库尔贝克-莱布勒 R^2。这种度量形式简洁，适用于广义线性模型，并可统一和拓展此前关于线性模型与非线性模型变量重要性的研究。文中还通过若干实例进行说明，所用数据来自公共卫生和保险领域。

关键词。对排列取平均；优势分析；拟合优度；层级划分；库尔贝克-莱布勒散度；回归

*巴塞尔大学，瑞士巴塞尔。电子邮件：sinan.acemoglu@unibas.ch
†巴塞尔大学，瑞士巴塞尔。电子邮件：christian.kleiber@unibas.ch
‡巴塞尔大学，瑞士巴塞尔。电子邮件：joerg.urban@unibas.ch

1 引言

在许多领域，评估回归分析中解释变量的重要性备受关注。相关文献跨越数十年，数量庞大且高度零散；众多领域都出现了平行发展和重新发现的现象，其中包括各种行为科学和社会科学、医学、生态学以及工商管理。以下综述颇具参考价值：Azen 和 Budescu (2003年)、Johnson 和 LeBreton(2004年)，以及Grömping (2007、2015)。

对于线性回归模型，一种重要的方法显然源自 Lindeman、Meranda 和 Gold（1980年），它通过将经典的 R^2 表示为模型中所有可能的回归变量子集所涉及的加权平均，来确定变量重要性。该方法当时并未引起太多关注，想必是因为计算负担相当沉重。（请注意，一个包含 10 个回归变量的模型——按当前标准来看规模相当适中——就需要进行 $2^{10} = 1024$次回归。）Lindeman 等人（1980年）提出的方法妙在易于解释，并且能够将整体拟合优度度量 R^2 的值“公平地”分解为各个回归变量的贡献。此后，人们使用不同术语多次重新发现并扩展了该方法，其中包括对排列取平均（下文将使用这一术语，参见 Kruskal，1984、1987）、层次划分（Chevan 和 Sutherland，1991）以及优势分析（Budescu，1993；Azen 和 Budescu，2006）。具体而言，层次划分和优势分析扩展了 Lindeman 等人（1980年）的思路，允许使用相当灵活的拟合优度度量来评估预测变量的贡献。它们的适用范围也不局限于线性回归。

Stufken （1992）指出，层次划分以及由此产生的对排列取平均，实质上是在使用Shapley值这一源自合作博弈论的概念，将回归的拟合度量进行分解。较早明确采用Shapley值框架的论文是Lipovetsky 和 Conklin （2001），但仍局限于线性回归。

近年来，统计计算的进步和计算能力的提升使得 Shapley 值的计算更加可行，而适用于非线性模型的灵活软件也促使人们开始寻找线性回归之外的变量重要性度量

。事实上，Chevan 和 Sutherland (1991年已经提出，可以通过分解 χ^2 统计量，运用层次划分来衡量经典非线性模型中预测变量的贡献。后来，Azen 和 Traxel(2009年将优势分析应用于逻辑回归，但采用的是某些伪R^2 度量。据我们所知，非线性模型的实证示例仍主要局限于二元响应模型，尽管一些软件实现提供了超出这一范围的功能，尤其是泊松回归。不过，某些提议及其相应实现存在问题，我们将在下文讨论。

本文旨在实现三个目标：(i) 阐述拟合优度度量应具备的理想性质，使 Shapley 值能够被解释为相对重要性乃至绝对重要性；(ii) 系统地将 Shapley 值应用于一类规模相当大的模型，包括广义线性模型及其某些扩展；(iii) 在使用 Shapley 值时，始终采用能够推广经典变体的拟合优度度量，同时将这一方法论拓展至更多情境。

我们表明，用于评估预测变量贡献的拟合优度度量必须谨慎选择，因为其性质对于解释所得的 Shapley 值至关重要。因此，我们提出了拟合优度度量应具备的理想性质，以帮助将所得的 Shapley 值解释为相对重要性，甚至是绝对重要性的度量。目前，似乎还没有一种普遍接受的程序用于评估广义线性模型（广义线性模型）中的变量重要性。因此，我们建议将 Shapley 值与满足特定要求且适用于所有广义线性模型的拟合度量结合使用，即 Kullback-Leibler R^2，下文记为 R^2_{KL}。该度量由 Cameron and Windmeijer （1997年）提出，与偏差和经典似然比统计量密切相关。具体而言，对于线性回归模型 R^2_{KL} 可化为经典的 R^2；而对于二元响应模型，它可化为 McFadden似然比指数，这或许是该领域所有伪R^2 拟合度量中最具代表性的一种。因此，我们的提议将 Shapley 值分解从线性回归模型扩展到范围大得多的广义线性模型（以及一些相关模型），从而为一系列非线性回归模型中的变量重要性提供统一方法。我们通过若干实证示例来说明这一方法论，所使用的数据

来自公共卫生和保险领域。

本文其余部分安排如下：第2 节概述 Shapley值的必要背景。第3 节讨论计算 Shapley 值时拟合优度度量应具备的理想性质。我们将在第 R^2_{KL}节证明其具备这些性质，该节为4。第5 节给出泊松回归、几何回归以及泊松障碍模型的示例；该模型具有两个线性预测变量，但其构件属于广义线性模型。这些示例也展示了选择拟合度量时需要权衡的因素。第6 节作总结。

2 回归中的 Shapley值

Shapley值是合作博弈论中的一种解概念，由 Shapley（1953）提出。Ferguson（2020）是了解博弈论术语及其背景的便捷参考资料。联盟形式的合作博弈（P, v）由一个有限的参与者集合 $P = \{1, \ldots, p\}$ 以及一个特征函数 v 描述： $2^P \to \mathbb{R}$ 该函数为幂集 2^P 的每个元素 S 赋予一个实数 $v(S)$。 $S \subseteq P$ 称为联盟， P称为大联盟，而 $v(S)$ 可解释为联盟 S 的成员作为一个整体行动时能够获得的收益。在合作博弈论中，特征函数 v 的一个标准条件是 $v(\varnothing) = 0$；也就是说，空集或联盟 $\varnothing$ 获得的收益为零。我们可以将其称为“零归一化”。

Shap 玩家 $i \in P$ 在给定特征函数 v i 下的 Shapley 值 $\varphi_i(P, v)$ s

$$\varphi_i(P, v) = \sum_{S \subseteq P \setminus \{i\}} \underbrace{\frac{|S|!\,(p - |S| - 1)!}{p!}}_{\text{权重}} \underbrace{\big(v(S \cup \{i\}) - v(S)\big)}_{\substack{\text{.....的边际贡献} \\ \text{玩家 } i\text{对联盟 } S\text{的}}}, \tag{1}$$

其中 $|S|$ 表示集合 S的基数。因此 $\varphi_i(P, v)$ 是玩家 i的平均边际贡献；该平均值是对 S 中 P 不包含玩家 i 的所有子集取平均得到的。按照 Kruskal (1987) 的说法，这种方法可以称为 对排列取平均。

在 Shapley值的各种性质中，效率性质 (Shapley, 1953)

$$\sum_{i \in P} \varphi_i(P, v) = v(P) \tag{2}$$

对本文尤为重要。它指出，在全体参与者组成的大联盟 P 上取值的特征函数值，等于所有 Shapley 值之和。

线性回归中的 Shapley 值

如上所述，Shapley 值通过对排序取平均这一原则，隐含地出现在变量重要性文献中。在回归语境下，“参与者”集合是用于预测特定结果变量的一组 $P = \{1, 2, \ldots, p\}$ 回归变量。这会产生由可形成的不同回归变量子集所定义的 2^p 个不同模型。特征函数 v 由适当的拟合优度度量发挥作用。在线性回归模型中，通常使用经典的 R^2，它满足 $v(\varnothing) = 0$。鉴于效率性质（2），每个 Shapley 值 $\varphi_i(P, v)$ 都可以解释为回归变量 i 对模型整体 R^2的贡献，因此在这一意义上，它度量了该回归变量的相对重要性。

Lindeman 等人(1980年，第 4.7 节）提出了一种基于半偏相关的变量重要性度量 $r^2_{(i \cdot S)^\circ}$。这些半偏相关 $r^2_{(i \cdot S)}$ 度量了响应变量 y 与回归变量 i之间的相关性，其中 $S \subseteq P$ 的其他预测变量的相关性已被剔除。该度量可以表示为纳入预测变量R^2 后所产生的 i增量的加权平均，具体而言，令 $v = R^2$

$$\begin{aligned} \varphi_i(P, R^2) &= \sum_{S \subseteq P \setminus \{i\}} \frac{|S|!\ (p - |S| - 1)!}{p!}\, r^2_{(i \cdot S)} \\ &= \sum_{S \subseteq P \setminus \{i\}} \frac{|S|!\ (p - |S| - 1)!}{p!} \left(R^2(S \cup \{i\}) - R^2(S)\right). \end{aligned} \tag{3}$$

此处 $R^2(S)$ 和 $R^2(S \cup \{i\})$ 分别对应于预测变量集合为 R^2的模型的 S 和 $S \cup \{i\}$。因此，Lindeman 等人 (1980年) 已隐含地提出在回归中使用 Shapley 值。

线性回归的应用不胜枚举；Grömping (2007, 2015) 提供了许多参考文献。

线性回归之外的 Shapley 值

Chevan 和 Sutherland (1991) 已建议将 Shapley 值用于线性回归模型以外的情形，并明确提及了逻辑回归、Probit 回归和对数线性回归。在层次划分框架下，Walsh、Papas、Crowther、Sim 和 Yoo (2004年) 使用对数似然作为特征函数，计算了 Logit 模型的 Shapley 值。在优势分析框架下，Azen 和 Traxel (2009年) 将其进一步应用于二元响应模型。为解决使用似然时固有的重新缩放问题，他们建议使用某些伪R^2 度量来计算 Shapley 值。在这些伪R^2 度量中，他们略微倾向于使用 McFadden (1973) 提出的似然比指数，以下记作 R^2_{McF}。较新的应用包括Yu、Zhou、Suh 和 Arcona (2015) 和 Lee 和 Dahinten (2021)。Nandintsetseg、Shinoda、Du 和 Munkhjargal(2018) 使用 R^2_{McF}进行泊松回归的优势分析；Tetteh、Ekem-Ferguson、Quarshie、Swaray、Ayanore、Seneadza、Asante 和 Yawson(2021) 也通过 Shapley 值评估泊松模型的变量重要性。

优势分析的应用甚至也存在于广义线性模型之外，参见 Shou 和 Smithson(2015)，其中给出了一个使用贝塔回归的示例。由于伪R^2 拟合度量（例如 McFadden似然比指数） R^2_{McF}最初是为二元数据设计的，因此可能不适用于其研究情形（该情形涉及连续分布）；他们使用了贝叶斯信息准则和似然比检验统计量等特征函数。此外，Lima、Ferreira 和 Leal (2021) 概述了一项涉及两部分模型（也称为门槛模型）的应用。他们的两部模型由两个广义线性模型构件组成：一个 Logit 模型和一个伽马回归模型。我们还在下文第5.2节给出一个两部模型的示例，探讨广泛使用的泊松障碍模型。

Shou 和 Smithson (2015) 强调，在非线性情形下，拟合优度度量的选择十分重要。事实上，我们将在第3节说明，谨慎选择拟合度量对于效率性质（2）的有效性至关重要，进而也关系到如何解释所得的 Shapley 值，以及如何统一相关方法论。

我们观察到 t 特征函数有许多潜在的选择

v。例如，机器学习（ML）文献近期呈现出一种趋势，即发展“可解释”或“可诠释”的机器学习，在这种方法中，Shapley 值被用于分解预测值或预测误差。该领域颇具影响力的论文包括 Štrumbelj 和 Kononenko (2010年）和 Lundberg 和 Lee （2017年），其中引用了计算机科学及相关领域的早期研究。在预测场景中，v 表示条件期望时，无论是 $v(\varnothing) = 0$，还是相对于增加更多预测变量的单调性，通常都不成立。因此，所得的基于 Shapley 值的分解并不代表拟合度量的分解。相比之下，与统计学文献中的早期发展一致，我们着重于分解拟合优度度量。这要求特征函数具有某些单调性属性。第3节将对此作进一步介绍。

软件

优势分析和层次划分已在多个 R 软件包中实现，可用于分解对数似然或伪R^2 度量（由 Azen 和 Traxel (2009年提出）。层次划分程序可通过**hier.part** 软件包获取（Mac Nally 和 Walsh，2004年），而优势分析则可通过 **dominanceanalysis** (Navarrete 和 Soares，2020年）软件包获取。对于线性模型，**relaimpo** (Grömping，2006年）软件包还提供了并非基于对排列取平均原理的方法。

3 拟合优度度量的作用

在实际应用中，变量重要性度量应当易于解释。本节将说明，在某些条件下，使用 Shapley 值可以实现这一目标。接下来，我们将讨论拟合优度度量（在原始博弈论背景下称为特征函数）应具备的理想性质。

3.1 单调性

从回归的角度来看，自然应要求特征函数在加入新的预测变量时单调不减，即 $v(S \cup \{i\}) \geq v(S)$ 对于任意 $S \subseteq P$ 和 $i \in P \setminus S$。这是一个有意义的条件，因为我们期望加入新的解释变量后，拟合优度度量会有所改善，或至少保持不变。根据（1），在此条件下得到的 Shapley 值均为非负。

3.2 下界

Shapley (1953) 要求特征函数满足 $v(\varnothing) = 0$，见第 2 节。在回归语境中，这一条件意味着，不使用任何回归变量（常数项除外）的模型拟合度为零。条件 $v(\varnothing) = 0$ 与单调性一起，还可推出 $v(S) \geq 0$ 对所有 $S \subseteq P$均成立。我们称之为下界条件。

在我们的语境中，Shapley 值 $\varphi_i(P, v)$ 表示对完整回归变量集上的拟合度量进行分解，$\sum_{i \in P} \varphi_i(P, v) = v(P)$。因此，（归一化后的）Shapley 值可以解释为相对于拟合模型 FM 的份额；也就是说，我们可以通过设定以下内容，为变量 i 定义一个重要性度量

$$impFM_i := \frac{\varphi_i(P,v)}{\sum_{j \in P} \varphi_j(P,v)} = \frac{\varphi_i(P,v)}{v(P)}, \qquad \text{with} \sum_{i=1}^{p} impFM_i = 1. \tag{4}$$

为便于后文使用，我们简要探讨当特征函数不满足零归一化条件时，对可解释性会产生什么影响。具体而言，考虑一个“伪特征函数” v^*，满足 $v^*(\varnothing) \neq 0$，并将由此得到的“伪 Shapley 值”记为 φ_i^*。接下来，定义 v 为 $v(\cdot) = v^*(\cdot) - v^*(\varnothing)$，它表示一个零归一化的特征函数。从方程（1）出发，我们可以看出，按定义，

$$\varphi_i^*(P, v^*) = \varphi_i(P, v), \tag{5}$$

因为伪 Shapley 值的构建基础是 v^* 的差值，而 Shapley 值的构建基础是 v 的差值。因此，我们也可以使用“伪 Shapley 值”对预测变量进行排序。具体而言，如果 $\varphi_A^* > \varphi_B^*$，那么在拟合模型中，预测变量 A 比预测变量 B 更重要。

此外，利用公式（5）和效率性质，我们有

$$\sum_i \varphi_i^*(P, v^*) \;=\; \sum_i \varphi_i(P, v) \;=\; v(P) \;=\; v^*(P) - v^*(\varnothing) \;\neq\; v^*(P). \tag{6}$$

因此，特征函数违反零归一化条件意味着，由此得到的伪 Shapley 值并不对应于预测变量对拟合模型整体 $v^*(P)$ 的贡献，因为 $\varphi_i^*(P, v^*)$ 不满足效率性质。事实上，所得伪 Shapley 值之和为 $v(P)$，因此应依据 $v(P)$ 而非 $v^*(P)$ 来解释。此外，基于伪 Shapley 值的量 $impFM_i$ 并不对应于拟合模型整体解释力 $v^*(P)$ 的相对贡献；也就是说，它们的总和不等于 1。这是方程 (6) 中偏移项 $v^*(\varnothing)$ 所导致的。因此，就解释而言，伪 Shapley 值的用途有限，只应用于预测变量的序数比较。我们将在第 5 节回到这一问题，并举例说明使用一种重要的伪特征函数——对数似然函数 ℓ——会引发的问题。对数似然 ℓ 通常不满足零归一化条件，因此会产生伪 Shapley 值，而非真正的 Shapley 值，从而带来各种解释上的问题。

3.3 上界

回想一下，空集对应于一个不包含回归变量的模型，因而描述的是“最差”模型。类似地，“最佳”模型可以通过数据驱动的方式定义，即为每个观测赋予其自身的回归系数。这在广义线性模型的背景下被称为 饱和模型 （例如 Dunn 和 Smyth，2018，第 274 页），对应于这样一种模型：对于给定的一组数据和给定的模型类型，其似然达到最大值（ McCullagh 和 Nelder，1989，第 33 页）。因此，在我们的语境中，它代表了另一个合适的参考点。

与上文讨论的“最差基准模型”一样，当使用无法相对于某个“最佳基准模型”进行解释的拟合优度度量 v 时，也可能出现解释问题。更形式化地说，假设 $v: S \to [0, b]$, $b \in \mathbb{R}_+$[1]

[1]由于单调性和零归一化条件，拟合优度的取值范围

且 $S \subseteq P'$。此处 P'表示饱和模型。那么， $v(\varnothing) = 0$和 $v(P') = b$分别表示“最差”模型和“最佳”模型的评估值。

在原始 Shapley(1953) 设定中，特征函数取值范围没有上界。然而，在回归语境中，如果没有上界，整体拟合度 $v(P)$ 就难以解释。这主要是因为“最佳”基准模型的拟合优度度量值 b 可能未知或无法获取（例如，在对数似然函数 ℓ的情况下）。在这里，我们仍可使用 $\varphi_i(P, v)$来评估相对重要性，方法与第3.1 节和3.2节类似。然而，可解释性方面的问题依然存在，因为一个看似较大的 Shapley值 $\varphi_i(P, v)$ 可能意味着变量非常重要，但模型的整体拟合度可能很差。只有在已知 b 的情况下，才能发现拟合不佳。换言之，较大的 Shapley值仍可能对应于相关性有限的预测变量。

如果已知“最佳基准模型”的拟合度，就可以解决这一问题。这样，Shapley值便可解释为相对于最佳模型的重要性， BM，

$$impBM_i := \frac{\varphi_i(P, v)}{b}. \tag{7}$$

如果我们还有一个特征函数，其有限上界等于 1，即 $b = 1$，那么现在甚至可以将 Shapley 值解释为相对于可实现的最佳模型而言的绝对重要性度量。此外，现在可以比较所有使用相同模型类和数据估计的模型。因此，为避免上述问题，我们建议使用上界为 1 的拟合优度度量。

3.4 拟合度量的结构可解释性

到目前为止，我们讨论了确保 Shapley 值能够被直接解释为重要性度量的条件，从而能够对回归变量进行排序，并评估它们的贡献相对于拟合模型或某个“最佳”模型是否足够大。

然而，除了前面几个小节提到的几点之外，对所得 Shapley 值的解释还高度取决于拟合度量本身的结构。

度量 v 是 $\mathbb{R}_+$ 的子集。

例如，线性回归模型中的经典 R^2 对应于模型所解释的总方差占比。因此，在 R^2 线性模型中使用经典度量时，某个预测变量的 Shapley值可以解释为该预测变量所解释的方差占比（Lindeman 等人，1980年）。相比之下，例如，使用基于似然的量作为拟合优度度量，通常不会在线性回归模型之外产生方差分解。

因此，尽管第3.1 –3.3节中的条件已允许在一定程度上解释 Shapley 值，但恰当地选择拟合优度度量还能带来更多洞见。因此，我们建议使用一种对于多种回归模型（如广义线性模型）都具有有意义的结构性解释的拟合优度度量。一个合适的候选是 Kullback-Leibler R^2，其性质总结于第4节。

3.5 拟合度量的理想性质

前面几小节阐述了将拟合优度度量选作特征函数的重要性、其结构、可解释性及其影响。鉴于上述问题，以下性质值得期望：

(i) 单调性：当加入新的预测变量时，v （弱）单调不减，

(ii) 下界：$v(\cdot) \geq 0$,

（iii）上界：v（饱和模型）$= 1$,

（iv）广义线性模型的 v 结构可解释性。

单调性条件 (i) 确保所有由此得到的 Shapley 值均为非负，并且可以通过相对排序进行比较。此外，条件 (ii) 确保 Shapley 值可以解释为对总体拟合度的贡献 $v(P)$，因此也可以解释为该变量相对于拟合模型的重要性。条件 (iii) 将这种解释进一步扩展到相对于可实现的最佳模型的重要性，即绝对重要性。表1总结了这些相互关系。

表 1：拟合度量的性质与相应 Shapley 值和伪 Shapley 值的性质之间的联系

	伪 Shapley 值作为相对变量重要性		Shapley 值作为相对变量重要性	Shapley 值作为相对绝对变量重要性
条件：贡献	非负	相对的排序	相对重要性对于拟合模型 $impFM$	相对重要性对于“最佳”模型 $impBM$
(i) 单调性	×	×	×	×
(ii) 下界			×	×
(iii) 上界				×

因此，如果满足条件 (i)–(iii)，所得的 Shapley 值便可以从相对重要性和绝对重要性的角度加以解释。在条件 (i)–(iii) 的基础上再满足条件 (iv)，还可确保从更深层的结构层面解释 Shapley 值。例如，当使用 R^2 作为拟合优度度量时，它们可以解释为线性回归中得到解释的方差比例。

此外，条件 (i)–(iii) 对一般的回归模型都有意义，并不局限于广义线性模型。条件 (iv) 则引出了本文的主要关注点：广义线性模型中的变量重要性。

4 Kullback-Leibler R^2 及其性质 s

在上一节中，我们强调，谨慎选择拟合优度度量至关重要，因为这既能避免误解结果，也有助于从更具结构性的层面解读 Shapley 值。

假设我们有一个来自 $f(y;\theta)$ 的随机样本 $y_1, \ldots, y_n$，其中该分布属于一个真正的（非弯曲）单参数指数族，并且

$$f(y;\theta) = \exp\{y\theta - b(\theta) + c(y)\},\ \theta \in \Theta, \tag{8}$$

其中，θ 是自然参数或规范参数，Θ 是实数轴上的一个区间，b是累积量函数，而 c 是不依赖于 θ的函数。对累积量函数 b 的不同选择会导出不同的模型。在广义线性模型中，f 的均值 μ 关于 θ 单调，且

通过线性预测变量 η_i 和已知的光滑可逆连接函数 g进行参数化，其中 $g(\mu_i) = x_i^\top \beta = \eta_i$; $x_i \in \mathbb{R}^p$ 是观测 i 的 p 回归变量向量，而$\beta \in \mathbb{R}^p$ 是回归系数向量。全文假设模型均包含常数项。采用最大似然法（ML）进行估计；最大似然估计量（MLE） θ 位于 Θ的内部。

在广义线性模型中，若采用 Kullback-Leibler R^2（记作 R^2_{KL}）作为拟合优度度量，就可以用统一的方法衡量变量重要性。 R^2_{KL} 由 Cameron and Windmeijer （1997年）提出。 R^2_{KL} 的主要优点是，它适用于任何基于（单参数）指数族的回归模型；概述请参见Cameron and Windmeijer （1997年）及其表 1。此外，许多广为人知的（伪） R^2 度量都是 R^2_{KL}的特例，其中包括线性回归模型中的经典 R^2 、二元响应模型中的 R^2_{McF} ，以及泊松回归中的偏差 R^2 （Cameron and Windmeijer，1996）。此外， R^2_{KL} 还可以根据似然比检验统计量来解释（见第 4.2节）。

4.1 Kullback-Leibler R^2

回顾一下，库尔贝克-莱布勒散度（Kullbackand Leibler，1951）定义为 s

$$K(\theta_1, \theta_2) := 2\, \mathsf{E}_{\theta_1} \left[\log \left(\frac{f(y, \theta_1)}{f(y, \theta_2)} \right) \right]. \tag{9}$$

它使用香农熵来衡量两个密度函数之间的信息差异，这两个密度函数在此分别由参数 θ_i、 $i = 1, 2$ 表示。E_{θ_1} 表示相对于由模型参数 θ_1 确定的模型的期望。对于单参数指数族，

$$K(\theta_1, \theta_2) = 2\, [(\theta_1 - \theta_2)\mu_1 - (b(\theta_1) - b(\theta_2))]. \tag{10}$$

由于 μ关于 f 的均值 θ是单调的，我们可以根据方便程度写作 $\mu = \mu(\theta)$ 或 $\theta = \theta(\mu)$, 也可以写作$K(\mu_1, \mu_2)$ 或 $K(\theta_1, \theta_2)$。在广义线性模型的设定中，如果 θ_1 表示具有 $\mu_1 = \mathbf{y}$ 的饱和模型，其中 $\mathbf{y}$ 表示数据，那么我们有

$$K(\mathbf{y}, \mu_2) = 2\, [(\theta(\mathbf{y}) - \theta_2)\mathbf{y} - (b(\theta(\mathbf{y})) - b(\theta_2))].$$

真正的单参数指数族是无曲率的或平坦的（Vos，1991），且库尔贝克-莱布勒散度满足毕达哥拉斯关系（Efron，1978；Hastie，1987），

$$K(\hat{\mu}, \mu) \;=\; K(\mathbf{y}, \mu) - K(\mathbf{y}, \hat{\mu}),$$

其中，$\hat{}\mu$ 表示拟合模型，μ 表示其他某个模型。

就我们的目的而言，$\mu = \hat{}\mu_0$（仅含常数项的模型）尤为重要。相应的库尔贝克-莱布勒 R^2 现在定义为

$$R^2_{\mathrm{KL}} \;=\; 1 - \frac{K(\mathbf{y}, \hat{\mu})}{K(\mathbf{y}, \hat{\mu}_0)} \quad \in [0, 1]. \tag{11}$$

其基本思想是，最简单的仅含常数项的模型——即线性模型中的 $\hat{}\mu_0 = \bar{y}\mathbf{1}_n \in \mathbb{R}^n$，其中 $\mathbf{1}_n$ 为由 1 组成的 n向量——与“最佳”模型的偏差最大，也就是说，它在预先指定的模型类中使库尔贝克-莱布勒散度达到最大。如果某个回归变量能为模型预测值 $\hat{}\mu$ 提供解释力并带来改进，那么我们有 $K(\mathbf{y}, \hat{}\mu) < K(\mathbf{y}, \hat{}\mu_0)$，从而得到更大的 R^2_{KL} 度量。第 3 节中的单调性条件按构造即得到满足，因为对于给定模型，$\ell(\mathbf{y}, \hat{}\mu)$ 的对数似然不会因替代模型中增加了一个预测变量而降低。结合下文的公式 (13) 和公式 (11)，库尔贝克-莱布勒 R^2 满足上下界条件，因为其取值介于 0 和 1 之间；具体而言，$R^2_{\mathrm{KL}} = 0$若 $\hat{}\mu = \hat{}\mu_0$，且 $R^2_{\mathrm{KL}} = 1$ 若 $\hat{}\mu = \mathbf{y}$。

Cameron and Windmeijer (1997年) 指出，库尔贝克-莱布勒散度也可以直观地解释为一种不确定性度量。具体而言，拟合模型相对于最优模型的“偏离”对应于经验不确定性，而经验不确定性可通过使用响应变量和 $\hat{}\mu$ 的库尔贝克-莱布勒散度来度量。因此，方程（4）和（7）中的重要性度量可以解释如下：

- 预测变量 i 的 Shapley值等于由预测变量 i解释的经验不确定性所占的比例 $\varphi_i = impBM_i$，因为 R^2_{KL} 的上界为 1。

- 相对于拟合模型的贡献 $impFM_i = \varphi_i / v(P)$ 等于由预测变量 i解释的已解释经验不确定性所占的比例。

总而言之，$v = R^2_{\mathrm{KL}}$ 是一种方便的拟合度量，因为它 (a) 具有第3节讨论的性质 (i)–(iii)，(b) 将经典的 R^2 度量推广到了

线性回归之外，(c) 基于信息论中的成熟概念，且 (d) 对于广泛使用的广义线性模型（GLM）模型类而言，形式简单。

下一小节将从似然比检验统计量的角度进一步阐释。

4.2 Kullback-Leibler R^2 与似然比统计量

在回归建模中，我们希望预测值向量为$\hat{\mu}$的拟合模型中的系数，相较于仅包含常数项、即预测值向量为$\hat{\mu}_0$的模型，整体上具有显著性。可以使用似然比（LR）统计量对此进行检验，

$$LR \;=\; 2\left(\ell(\mathbf{y},\hat{\mu}) - \ell(\mathbf{y},\hat{\mu}_0)\right). \tag{12}$$

利用指数族的 Hoeffding 表示（Efron，1978；Hastie，1987），Kullback-Leibler散度也可以用似然来表示，

$$K(\mathbf{y},\hat{\mu}) \;=\; 2\left(\ell(\mathbf{y},\mathbf{y}) - \ell(\mathbf{y},\hat{\mu})\right), \tag{13}$$

其中，$\ell(\mathbf{y},\mathbf{y})$是饱和模型的对数似然，而 $\ell(\mathbf{y},\hat{\mu})$是由$\hat{\mu} \in \mathbb{R}^n$表示的拟合模型的对数似然。

在广义线性模型术语中，方程（13）中的量就是（残差）偏差；对于广义线性模型，它的作用相当于经典线性模型中的残差平方和。因此，对于广义线性模型，Kullback-Leibler R^2 与偏差解释比例完全相同。

从方程（13）、（11）和（12）可以明显看出，R^2_{KL} 和 LR 密切相关。事实上，在 Kullback-Leibler R^2提出之前，Magee (1990) 就已提出通过经典的基于似然的检验统计量来定义拟合度量。具体而言，R^2_{KL} 是似然比统计量的一个标量倍数（Cameron and Windmeijer，1997年），

$$R^2_{\mathrm{KL}} \;=\; \frac{1}{K(\mathbf{y},\hat{\mu}_0)}\, LR. \tag{14}$$

鉴于效率性质，我们有 $\sum_i \varphi_i(P,\; R^2_{\mathrm{KL}}) \;=\; R^2_{\mathrm{KL}}$，因此 Shapley 值也对应于缩放似然比检验统计量的某种分解，

$$K(\mathbf{y},\hat{\mu}_0) \sum_i \varphi_i(P, R^2_{\mathrm{KL}}) = LR. \tag{15}$$

由于所有 Shapley 值都乘以同一个常数 $K(\mathbf{y}, \hat{}\mu_0)$（即零偏差），这带来了进一步的解释：预测变量 i 的 Shapley值可以解释为该预测变量对模型整体似然比统计量的贡献，差别仅在于一个常数因子。

4.3 Kullback-Leibler R^2 和McFadden似然比指数

如上所述，对于二元响应模型，广泛使用的一种伪R^2 度量是McFadden似然比指数

$$R^2_{\mathrm{McF}} \;=\; 1 - \frac{\ell(\mathbf{y}, \hat{\mu})}{\ell(\mathbf{y}, \hat{\mu}_0)}. \tag{16}$$

利用方程（13）和（11），可得

$$R^2_{\mathrm{McF}} \;=\; \left(1 - \frac{\ell(\mathbf{y}, \mathbf{y})}{\ell(\mathbf{y}, \hat{\mu}_0)}\right) R^2_{\mathrm{KL}} = \zeta\, R^2_{\mathrm{KL}}. \tag{17}$$

对于伯努利分布，其中 $y_i \in \{0, 1\}$，

$$\ell(\mathbf{y}, \mathbf{y}) \;=\; \sum_{i=1}^{n} \left(y_i \log(y_i) + (1 - y_i) \log(1 - y_i)\right) \;=\; 0, \tag{18}$$

这意味着 $\zeta = 1$；因此，在二元响应情形下有 $R^2_{\mathrm{McF}} = R^2_{\mathrm{KL}}$ 。然而，在该情形之外，通常有 $\zeta \neq 1$，这导致 R^2_{McF} 违反上界条件。下一节将用泊松模型说明这一问题。

5 示例

本节通过示例介绍如何使用 Shapley 值评估所选广义线性模型（及其某些扩展）的变量重要性。随后，我们用 $\ell(S)$ 和 $R^2_{\mathrm{KL}}(S)$ 表示包含预测变量 $S \subseteq P$ 的回归模型所计算的拟合优度度量，其中 P 的基数等于 p。例如，$\ell(\varnothing)$ 对应于仅包含常数项的模型的对数似然。类似地，$\hat{}\mu_S$ 是使用预测变量集合 S 的回归模型计算得出的预测值向量。

以下几点值得强调：

1. 简介从相对重要性和绝对重要性的角度解读 Shapley 值 .

2. 违反上下界条件的后果。

3. 基于不同似然量推导出的 Shapley 值之间的关系。

5.1 泊松回归

我们从泊松回归开始，其基于概率密度

$$f(y_i; x_i, \mu_i) \;=\; \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \quad \text{for } y_i \in \mathbb{N}_0, \tag{19}$$

其中 μ_i =E($y_i \mid x_i$)，且 η_i =log(μ_i)。

我们使用健康经济学领域的数据，这些数据最初由 Cameron and Trivedi（1986年）分析，另见 Cameron and Trivedi（2013年），可从 Journal of Applied Econometrics[2] 的数据档案中获取。R 用户也可以从 R 软件包 AER（Kleiber and Zeileis，2008年）中以 DoctorVisits 为名获取这些数据，其中详细介绍了所有变量。响应变量是就诊次数（visits），最多计数为 9。回归变量提供了样本中患者健康状况及社会经济特征的信息。具体而言，我们使用年龄（age）、性别（gender）、健康（health）、疾病（illness）、收入（income）、慢性病（lchronic）、nchronic、private 和 reduced 这些变量。采用这组回归变量的模型根图（Kleiber and Zeileis，2016年）证实，泊松模型适用于这些数据；该图见图 1。

The Shapley 值通过 Kullback-Leibler R^2计算，定义如下：

$$R^2_{\mathrm{KL,Poi}} \;=\; 1 - \frac{\sum_{i=1}^{n}\left[y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)\right]}{\sum_{i=1}^{n} y_i \log(y_i/\bar{y})}, \tag{20}$$

即 Cameron and Windmeijer R^2 提出的泊松偏差（1996）。为说明违反第3 节所讨论要求的后果，我们还使用 McFadden 的 R^2。例如，后者在R 软件包 **dominance-analysis** 中实现，可与多种广义线性模型配合使用，包括泊松回归。

根据方程（17），我们知道，对于所有满足 $\ell(\mathbf{y},\mathbf{y}) \neq 0$ 的回归模型，也就是说，$\zeta \neq 1$，都有 $R^2_{\mathrm{McF}} \neq R^2_{\mathrm{KL}}$。对于我们的数据和所选的泊松模型，有 $\zeta = 0.71$,

[2]http://qed.econ.queensu.ca/jae/1997-v12.3/mullahy/

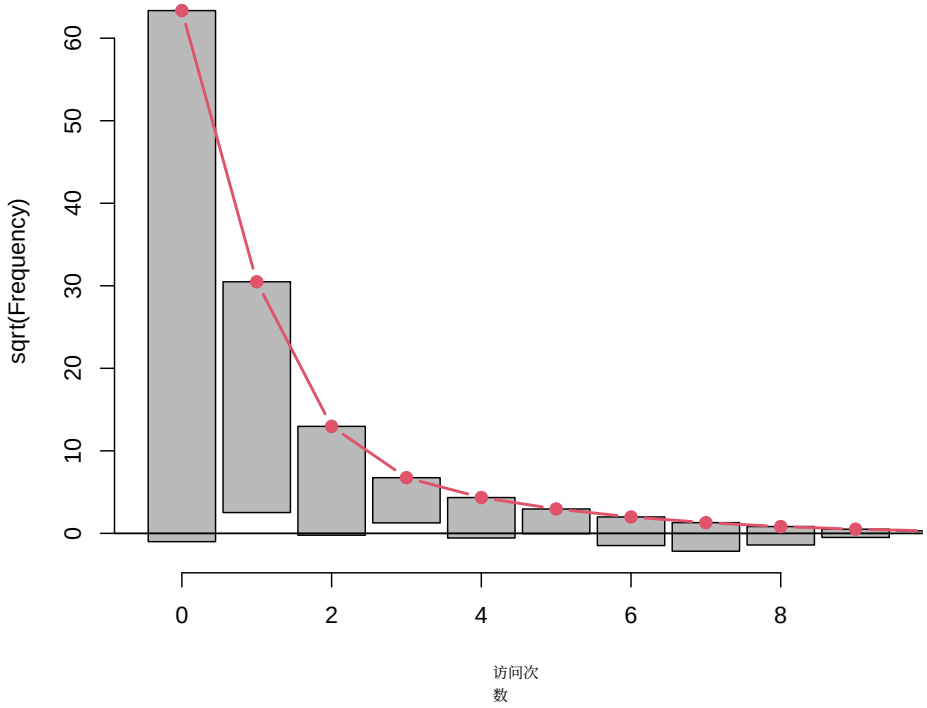

图1：使用所有回归变量的泊松回归模型根图。

因此， R^2_{McF} 的取值范围小于 R^2_{KL}的取值范围，而后者是单位区间。这导致该泊松模型的 R^2_{McF} 违反了上界条件。附录 A 表明，对于 Shapley 值，我们也有 $\varphi_i(P, R^2_{\text{McF}}) = \zeta\, \varphi_i(P, R^2_{\text{KL}})$。[3]因此，基于 R^2_{McF} 的 Shapley 值不再能相对于最佳模型进行解释，除非将其按 ζ重新缩放。

我们在表2中说明了这些问题。模型解释的经验不确定性约为22.11%。最大的贡献来自预测变量reduced；其相对重要性 $impFM_{\texttt{reduced}}$约为56.54%。此外，由于饱和模型的 R^2_{KL}拟合优度为 1，因此 Shapley 值也可以解释为相对于最佳模型的重要性度量（参见第3.3节）。因此，从 $\varphi_{\texttt{reduced}}(P, R^2_{\text{KL}}) = impBM_{\texttt{reduced}} = 12.5\%$ 可以看出，预测变量reduced解释了总经验不确定性的12.5% ，因此对于解释就诊次数

[3]Any 表中出现的数值偏差2 是由于四舍五入造成的。

表 2：使用 $v = R^2_{\mathrm{KL}}$ 和
$v = R^2_{\mathrm{McF}}$，分别。

	减少	疾病	健康	慢性病	age	$v(P)$
$v = R^2_{\mathrm{KL}}$	0.1250	0.0415	0.0207	0.0111	0.0110	0.2211
$v = R^2_{\mathrm{McF}}$	0.0884	0.0293	0.0146	0.0079	0.0078	0.1564

具有相当重要的意义。相比之下，使用 $v = R^2_{\mathrm{McF}}$ 进行的重要性评估无法支持此类解释，除非明确计算 ζ，因为如上所述，在此应用中，R^2_{McF} 违反了上界条件。这说明选择拟合优度度量时应格外谨慎，因为其选择会影响重要性度量的可解释性。

我们还要补充说明，由于 R^2_{KL} 对应于缩放似然比检验统计量，因此来自 R^2_{KL} 的 Shapley 值也可以解释为对该检验统计量的贡献。在我们的案例中，reduced 是贡献最大的预测变量。

5.2 泊松障碍回归

在计数数据回归领域，许多数据集都受到大量零观测值的困扰，因此泊松模型等经典模型无法提供适当的拟合。一种更灵活的设定是障碍回归模型，也称为两部模型，最初由 Mullahy (1986年) 提出。更正式地说，障碍模型是两个模型的组合，即 f_1 和 f_2，其中 f_1 表示二元响应部分，通常采用 logit 形式，而 f_2 是一个在 $y = 1$处进行左截断的计数数据模型。总体而言，障碍模型表示为

$$f(y_i; x_{i1}, x_{i2}, \tau_1, \tau_2) = \begin{cases} f_1(0; x_{i1}, \tau_1), & \text{for } y_i = 0, \\ \dfrac{1 - f_1(0; x_{i1}, \tau_1)}{1 - f_2(0; x_{i2}, \tau_2)} f_2(y_i; x_{i2}, \tau_2), & \text{for } y_i > 0, \end{cases} \tag{21}$$

其中，x_{i1}，x_{i2} 分别是观测 i 的回归变量向量，τ_1，τ_2 分别是相应的回归系数向量。一般而言，$x_{i1} \neq x_{i2}$，但以下设定

$x_{i1} = x_{i2}$ 在实证文献中相当常见。

障碍模型的对数似然 ℓ_{hurdle}可拆分为两个部分（Mul-lahy，1986年）：二元响应模型的对数似然 ℓ_{binary}以及零截断计数回归模型的对数似然 $\ell_{\text{zt-count}}$；即 $\ell_{\text{hurdle}} = \ell_{\text{binary}} + \ell_{\text{zt-count}}$。这意味着，可以分别拟合这两个部分来估计障碍模型，也可以逐部分评估变量重要性。

对于二元部分，我们使用 Logit 模型，其中 y_i^* 表示正计数的指示变量，

$$f(y_i^*; x_{i1}, \mu_{i1}) = \mu_{i1}^{y_i^*} (1 - \mu_{i1})^{(1-y_i^*)}, \quad \text{for } y_i^* \in \{0, 1\}, \tag{22}$$

其中 $\mu_{i1} = \mathsf{P}(y_i^* = 1 \mid x_{i1}) = \mathsf{E}(y_i^* \mid x_{i1})$ 且 $\eta_i = \log(\mu_{i1}/(1 - \mu_{i1}))$。

对于正计数，我们使用零截断泊松模型，其中

$$f(y_i; x_{i2}, \mu_{i2}) = \frac{\mu_{i2}^{y_i}}{(e^{\mu_{i2}} - 1)\, y_i!}, \quad \text{for } y_i \in \mathbb{N}, \tag{23}$$

其中 $\mu_{i2} = \mathsf{E}(y_i \mid x_{i2})$ 且 $\eta_i = \log(\mu_{i2})$。零截断泊松分布仍属于指数族，因此自然可导出一个广义线性模型。

因此，我们模型的两个对数似然组成部分为

$$\begin{aligned} \ell_{\text{logit}}(\mathbf{y}^*, \mu_1) &= \sum_{i=1}^{n} \left[y_i^* \log(\mu_{i1}) + (1 - y_i^*) \log(1 - \mu_{i1}) \right], \\ \ell_{\text{ztPoi}}(\mathbf{y}, \mu_2) &= \sum_{\{i: y_i > 0\}} \left[y_i \log(\mu_{i2}) - \log\left(e^{\mu_{i2}} - 1\right) - \log(y_i!) \right]. \end{aligned} \tag{24}$$

总的来说，方程（24）表明，泊松障碍模型包含两个广义线性模型构件：用于二元部分的 Logit 模型，以及用于正计数的零截断泊松回归模型。它们还表明，可以分别对这两个构件使用 Shapley值方法来评估变量重要性。

我们使用这个泊松障碍模型来拟合Yip and Yau (2005年曾分析过的汽车保险数据。这些数据可通过`AutoClaim` 数据获取，该数据来自R软件包**cplm**(Zhang，2013年）。具体而言，可通过一个二元因子提取相关观测子集：`IN_YY`，该因子表示观测是否纳入 Yip and Yau 的论文。我们关注的响应变量是过去五年的索赔数量，`cfreq5`，其最大计数为5。Yip and Yau 在主要分析中使用了五个回归变量，但他们最初使用的是包含 13 个回归变量的较大数据集，而这些变量又取自一个更大的数据集。鉴于我们关注

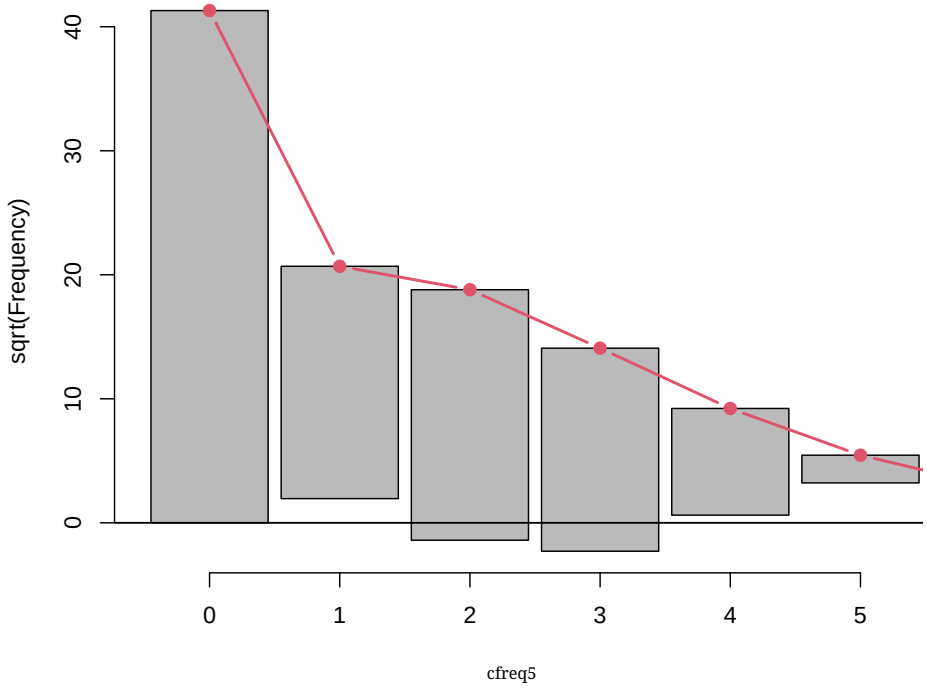

图2：使用全部13个回归变量的泊松障碍模型根图。

变量重要性，我们采用他们最初的 13 个回归变量：年龄、地区、车型、教育程度、性别、收入、工作类别、婚姻状况、红色、已撤销、单亲、用途和违规。[4] 另请参阅Yip and Yau (2005年）以及`AutoClaim` 数据的文档，以了解有关这些预测变量的更多信息。图2 中的根图证实，使用这些预测变量的泊松障碍模型能够充分拟合索赔频率变量。

零截断泊松模型的 Kullback-Leibler R^2 $R^2_{\text{KL, ztPoi}}$，定义为

$$R^2_{\text{KL, ztPoi}} = 1 - \frac{\sum_{\{i:y_i>0\}} \left[y_i \log(y_i/\hat{\mu}_i) - \log(\exp(y_i)-1) + \log(\exp(\hat{\mu}_i)-1)\right]}{\sum_{\{i:y_i>0\}} \left[y_i \log(y_i/\bar{y}_+) - \log(\exp(y_i)-1) + \log(\exp(\bar{y}_+)-1)\right]}, \tag{25}$$

其中 $\bar{y}_+ = n_+^{-1} \sum_{\{i:\ y_i>0\}} y_i$，$n_+$ 对应于正响应的数量

[4] AutoClaim数据中变量的原始名称按顺序分别为：AGE、AREA、CAR_TYPE、MAX_EDUC、GENDER、INCOME/10000、JOBCLASS、MARRIED、RED_CAR、REVOLKED、PARENT1、CAR_USE 和 MVR_PTS。请注意，收入是将 INCOME 除以 10000 得到的。响应变量 cfreq5 的原始名称为 CLM_FREQ5。

($y_i > 0$)。对于 t 使用 logit 链接函数的二元部分（CameronandWindmeijer，1997年 ）

$$R^2_{\text{KL, logit}} = 1 - \frac{\sum_{i=1}^{n} \left[y_i^* \log(\hat{\mu}_i) + (1 - y_i^*) \log(1 - \hat{\mu}_i) \right]}{\sum_{i=1}^{n} \left[y_i^* \log(\bar{y}^*) + (1 - y_i^*) \log(1 - \bar{y}^*) \right]}. \tag{26}$$

如上所述，二元部分的 Kullback-Leibler R^2 与 R^2_{McF} 相同。表 3：使用 $v = R^2_{\text{KL}}$、$v = \ell - \ell(\varnothing)$ 和 $v^* = \ell$ 得出的障碍模型正值部分中最大的五个 Shapley 值和伪 Shapley 值。

	车型	工作类别	red	educ	单亲	$v(P)$ 或 $v^*(P)$
$v = R^2_{\text{KL}}$	0.0103	0.0067	0.0045	0.0036	0.0036	0.0340
$v = \ell - \ell(\varnothing)$	2.9199	1.9136	1.2756	1.0252	1.0212	9.6376
$v^* = \ell$	2.9199	1.9136	1.2756	1.0252	1.0212	-1453.3342

表 3 列出了使用全部 13 个变量时，障碍模型正值部分中最大的五个 Shapley 值，以及由此得到的拟合优度度量 R^2_{KL}、对数似然和平移对数似然。（选择这些拟合度量，是因为它们可在软件实现中使用，具体来说，可见于 R 包 hier.part 和 dominanceanalysis。）该表清楚地展示了违反上下界条件所引发的问题，正如第 3.2 节和第 3.3 节所述。首先，与对数似然相关的伪 Shapley 值（表 3 底行）可能会让人认为这些变量具有一定的相关性，尽管就 R^2_{KL} 而言，总体拟合度相当差，因为模型仅解释了约 3.4% 的经验不确定性。相应的总体对数似然值 $v^*(P) = -1453.33$ 在这里并未提供多少有用信息：它是负值，且难以解释。其次，平移对数似然 $v = \ell - \ell(\varnothing)$ 满足效率条件，相应的总体值 $v(P) = 9.64$ 可能会让人认为模型表现良好（请记住，值为零对应最差的模型）。然而，其可解释性仍然不足，因为最佳模型的拟合度未知或未经计算。缺少最佳模型这一参照点，便无法判断 $v(P) = 9.64$ 是“较小”还是“较大”。对于满足上下界条件的拟合度量，例如 R^2_{KL}，则不会出现这个问题，在这种情况下，

值 $v(P) = 0.034$ 可与内置上界 1 进行比较。使用 R^2_{KL} 时，Shapley 值除了可以从相对重要性的角度进行解释，也可以从绝对重要性的角度进行解释。例如，车型的重要性约为工作类别的 1.54 倍，但从绝对重要性来看，这两个预测变量都不重要。

表 4：分别使用 $v = R^2_{\text{KL}}$、$v = \ell - \ell(\varnothing)$ 和 $v^* = \ell$ 时，障碍模型二元部分中最大的五个 Shapley 值和伪 Shapley 值。

	违规	area	车型 工作类别		educ	$v(P)$ 或 $v^*(P)$
$v = R^2_{\text{KL}}$	0.1612	0.0580	0.0038	0.0034	0.0024	0.2353
$v = \ell - \ell(\varnothing)$	303.7332	109.3837	7.1795	6.4305	4.4890	443.5264
$v^* = \ell$	303.7332	109.3837	7.1795	6.4305	4.4890	-1441.0973

二元部分的结果汇总于表4中。与之前一样，违反上下界条件会导致解释问题。对于 $v^* = \ell$，可以根据相对重要性对预测变量进行排序，但在不知道$impFM$）和“最佳”（$impBM$）模型的情况下，无法陈述其相对于拟合 $\ell(\varnothing)$ 的重要性。对于 $v = \ell - \ell(\varnothing)$，关于 $impFM$ 的陈述则成为可能。违规的 Shapley 值和地区的 Shapley 值最大——这些回归变量分别解释了拟合模型的 68.5% 和 24.7%——而其他预测变量的重要性似乎要低得多。使用 $v = R^2_{\text{KL}}$ 可以提供进一步的洞见，因为此时 Shapley 值可以解释为绝对重要性度量。相对于最佳模型，预测变量违规和地区分别解释了 16.12% 和 5.80%。

比较表 3 和 4 可以发现另一个有趣的细节：二元响应模型解释了 23.53%的经验不确定性，而表3 中的零截断泊松模型仅解释了 3.40%。因此，这个两部模型的二元部分表现远胜于零截断泊松部分。实际而言，这意味着现有回归变量有助于对索赔发生率建模，但对于对索赔数量进行精确建模，其相关性有限。

此外，由于 R^2_{KL} 是基于似然的拟合优度度量，因此由使用所导致的‘扭曲’

（伪）Shapley 值 $v = \ell - \ell(\varnothing)$ 和 $v^* = \ell$ 可以进行量化。这三种拟合优度度量呈线性关系，具体而言，可以证明（见附录A）

$$\varphi_i(P, R^2_{\mathrm{KL}}) \;=\; \underbrace{\frac{1}{\ell(\mathbf{y},\mathbf{y}) - \ell(\mathbf{y},\hat{\mu}_0)}}_{=:C}\, \varphi^*_i(P,\ell) \;=\; \underbrace{\frac{1}{\ell(\mathbf{y},\mathbf{y}) - \ell(\mathbf{y},\hat{\mu}_0)}}_{=:C}\, \varphi_i(P, \ell - \ell(\varnothing)), \qquad (27)$$

其中， C 为零偏差。换言之，由此得到的（伪）Shapley 值继承了拟合度量之间存在的线性关系。对于当前数据，零截断泊松模型和 Logit 模型分别有$C^{-1} = 283.75$ 和 $C^{-1} = 1884.62$ 。[5]

5.3 几何回归

计数数据的另一个典型问题是过度离散，即数据集中的变异程度超过了基于给定均值模型所预期的程度（其隐含的参照点是泊松模型）。因此，我们的最后一个示例介绍一种允许存在大量过度离散的计数数据模型。同样，我们使用健康经济学数据，这些数据最初由 Deb 和 Trivedi（1997年）进行分析，另见 Cameron and Trivedi（2013年）。这些数据也可从 Journal of Applied Econometrics 的数据档案中获取[6]。对于 R 用户而言，这些数据还以 NMES1988 的名称收录于 R 软件包 AER（Kleiber and Zeileis, 2008年）中，其中提供了所有变量的详细说明。响应变量是医生诊室就诊次数（visits），最大计数为 89。回归变量 adl、afam、age、chronic、employed、gender、health、income、insurance 和 married 提供了有关样本个体健康状况和社会经济地位的信息。

图 3 展示了使用所有回归变量的模型根图，证实几何回归模型适用于这些数据。因此，我们采用

$$f(y_i; x_i, \mu_i) \;=\; \frac{\mu_i^{y_i}}{(1+\mu_i)^{(y_i+1)}}, \quad \text{for } y_i \in \mathbb{N}_0, \qquad (28)$$

[5]根据表格计算得出的数值 C 略有不同，这是由于舍入造成的。[6] http://qed.econ.queensu.ca/jae/1997-v12.3/deb-trivedi/

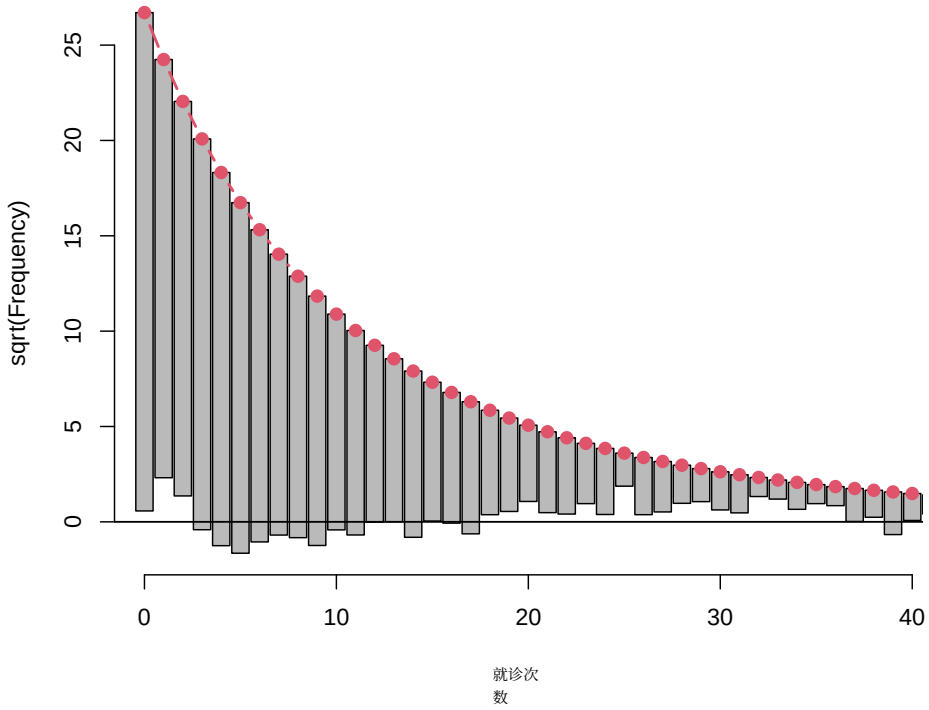

图3：使用所有回归变量的几何回归模型根图。

其中 μ_i =E($y_i \mid x_i$) 且 η_i = log(μ_i)。与计数数据的常见做法一样，这里使用对数连接函数，尽管它并非该模型的典范连接函数。

Shapley 值通过以下方法获得 R^2_{KL}，此处形式为 (Cameronand Windmeijer，1997 年)

$$R^2_{\mathrm{KL,geo}} = 1 - \frac{\sum_{i=1}^{n} \left[y_i \log\left(y_i/\hat{\mu}_i\right) - (1+y_i) \log\left((1+y_i)/(1+\hat{\mu}_i)\right) \right]}{\sum_{i=1}^{n} \left[y_i \log\left(y_i/\bar{y}\right) - (1+y_i) \log\left((1+y_i)/(1+\bar{y})\right) \right]}. \tag{29}$$

表5： $v = R^2_{\mathrm{KL}}$的几何回归模型中最大的五个 Shapley 值。

	慢性病	健康	保险	adl	afam	$v(P)$
$v = R^2_{\mathrm{KL}}$	0.0535	0.0233	0.0086	0.0056	0.0017	0.0953

表5 给出了基于 R^2_{KL}的 Shapley 值。我们不将其与

其他拟合度量进行比较，因为我们不知道是否有实证研究将变量重要性度量与几何回归结合使用。预测变量慢性病（对应慢性疾病的数量）贡献最大，其次是健康。然而，尽管慢性病在拟合模型中的相对重要性较高，达到 $impFM_{\texttt{chronic}} = 56.14\%$，完整模型的解释力却并不突出，其$R^2_{\mathrm{KL}}$ 约等于9。53%。慢性病的绝对重要性约为5。35%。因此，回归变量慢性病仅能解释总经验不确定性的5。35% ，因此对于解释就诊次数而言，其解释力似乎有限。这个例子再次凸显了具有第3节所述性质的拟合优度度量的实用性，因为它们可以通过绝对重要性来解读。

6 结论

在许多领域，理解解释变量在回归模型中的重要性都至关重要。一种广受欢迎的方法以Shapley值为基础，这一概念源自博弈论。计算 Shapley值的一个关键组成部分是特征函数，或者用回归术语来说，是一种合适的拟合优度度量。在统计学文献中，这一思想主要应用于线性回归模型，而经典的 R^2 是自然的起点。在这一背景下，Shapley 值为经典的 R^2提供了一种“公平”的分解。

然而，目前尚无广受认可的框架可用于评估广义线性模型中的变量重要性。我们在针对线性回归和二元响应模型的既有研究基础上，提出了一种适用于广义线性模型的统一方法。我们还提出并讨论了拟合优度度量应具备的理想性质，其中有些性质也适用于广义线性模型之外的回归模型。我们证明，这些性质使得 Shapley 值能够被解释为相对重要性和绝对重要性的度量。此外，我们建议采用 Kullback-Leibler R^2；对于广义线性模型，它与偏差解释比例相同，并且推广了若干广为人知的拟合度量，例如经典的 R^2 和用于二元响应模型的 McFadden似然比指数。

Kullback-Leibler R^2 也可能是若干

并非基于构成指数族的分布的非线性回归模型所选用的拟合优度度量。目前对此仍在研究中。不过，本文通过在第5节使用泊松障碍模型，迈出了超越广义线性模型框架的第一步。该模型包含广义线性模型构件，但其两部分结构使其比广义线性模型更灵活。

参考文献

Azen, R. 和 Budescu, D. V. (2003), ‘用于比较多元回归中预测变量的优势分析方法’, 《心理学方法》 8(2), 129–148.

Azen, R. 和 Budescu, D. V. (2006), ‘多变量回归模型中的预测变量比较：优势分析的扩展’, 《教育与行为统计学杂志》31(2), 157–180.

Azen, R. 和 Traxel, N. (2009), ‘运用优势分析确定逻辑回归中预测变量的重要性’, 《教育与行为统计学杂志》34(3),319– 347.

Budescu, D. V. (1993), ‘优势分析：一种解决多元回归中预测变量相对重要性问题的新方法’, 《心理学通报》 114(3), 542–551.

Cameron, A. C. 和 Trivedi, P. K. (1986), ‘基于计数数据的计量经济模型：若干估计量和检验的比较与应用’, *Journal of Applied Econometrics*1(1), 29–53.

Cameron, A. C. 和 Trivedi, P. K.（2013年）， 《计数数据回归分析》， 第2版， 剑桥大学出版社， 剑桥。

Cameron, A. C. 和 Windmeijer, F. A. G.（1996）， ‘用于计数数据回归模型的 R-平方度量及其在医疗保健利用中的应用’， 《style id='2'>《商业与经济统计学杂志》14（2）， 209–220。

Cameron, A. C. 和 Windmeijer, F. A. G.（1997）， ‘用于某些常见非线性回归模型的 R 平方拟合优度度量’， 《计量经济学杂志》77(2), 329–342。

Chevan, A. 和 Sutherland, M.（1991），‘层次划分’，美国统计学家45(2), 90–96。Deb, P. 和 Trivedi, P. K.（1997），‘老年人对医疗保健的需求：有限混合方法’，*Journal of Applied Econometrics*12(3), 313–336。Dunn, P. K. 和 Smyth, G. K.（2018），《广义线性模型及其 *R* 示例》*R*，施普林格，纽约。Efron, B.（1978），‘指数族的几何’，统计年刊6(2), 362–376。Ferguson, T. S.（2020），博弈论教程，世界科学出版社，新加坡。Grömping, U.（2006），‘R 中线性回归的相对重要性：re-laimpo 软件包’，统计软件杂志17(1), 1–27。Grömping, U.（2007），‘基于方差分解的线性回归相对重要性估计量’，美国统计学家61(2), 139–147。Grömping, U.（2015），‘回归模型中的变量重要性’，《威利跨学科评论：计算统计学》7(2), 137–152。Hastie, T.（1987），‘深入考察偏差’，美国统计学家41(1), 16–20。Johnson, J. W. 和 LeBreton, J. M.（2004），‘相对重要性指数在组织研究中的历史与应用’，组织研究方法7(3), 238–257。Kleiber, C. 和 Zeileis, A.（2008），《使用 *R* 进行应用计量经济学》*R*，使用R！，施普林格，纽约。Kleiber, C. 和 Zeileis, A.（2016），‘使用根图可视化计数数据回归’，美国统计学家70(3), 296–303。Kruskal, W.（1984），‘相对重要性的概念’，*Qüestiió* 8(1), 39–45。Kruskal, W.（1987），‘通过对排列取平均计算相对重要性’，美国统计学家41(1), 6–10。更正：第 41 卷（1987年），第 341 页。

Kullback, S. 和 Leibler, R. A.（1951），《论信息与充分性》，数学统计年刊**22(1)**，79–86。Lee, S. E. 和 Dahinten, V. S.（2021），《运用优势分析确定预测患者安全的安全文化最重要维度》，《国际环境研究与公共卫生杂志》**18(15)**。Lima, F.、Ferreira, P. 和 Leal, V.（2021），《健康与住房能源支出：一种两部模型方法》，*Processes* **9(6)**，943。Lindeman, R. H.、Meranda, P. F. 和 Gold, R. Z.（1980年），《双变量与多变量分析导论》，Scott, Foresman and Company，格伦维尤，伊利诺伊州。Lipovetsky, S. 和 Conklin, M.（2001），《博弈论方法下的回归分析》，《商业与工业中的应用随机模型》**17(4)**，319–330。Lundberg, S. M. 和 Lee, S.-I.（2017年），《解释模型预测的统一方法》，神经信息处理系统进展 **30**。Mac Nally, R. 和 Walsh, C. J.（2004年），《层次划分公有领域软件》，《生物多样性与保护》 **13**，659–660。Magee, L.（1990），《基于 Wald 和似然比联合显著性检验的R^2 度量》，美国统计学家 **44(3)**，250–253。McCullagh, P. 和 Nelder, J. A.（1989），《广义线性模型》，第2版，Chapman & Hall，伦敦/纽约。McFadden, D.（1973），定性选择行为的条件Logit分析，收录于P. Zarembka 编，《计量经济学前沿》，Academic Press，纽约，第105–142页。Mullahy, J.（1986），《某些修正计数数据模型的设定与检验》，《计量经济学杂志》**33(3)**，341–365。Nandintsetseg, B.、Shinoda, M.、Du, C. 和 Munkhjargal, E.（2018），《欧亚草原的寒季灾害：气候驱动还是人为因素》，《科学报告》 **8(1)**。

Navarrete, B. C. 和 Soares, C. F.（2020年），*dominanceanalysis*：优势分析。R 软件包版本 2.0.0。

Shapley, L. S.（1953），n人博弈的一个值，载于H. W. Kuhn 和 A. W. Tucker 编，《博弈论贡献，第二卷》，数学研究年刊，普林斯顿大学出版社，普林斯顿，第 307–317 页。

Shou, Y. 和 Smithson, M.（2015），《评估离散度预测变量：优势分析与贝叶斯模型平均的比较》，*Psychometrika*80(1)，236–256。

Štrumbelj, E. 和 Kononenko, I.（2010），《使用博弈论高效解释个体分类》，机器学习研究杂志 11，1–18。

Stufken, J.（1992），《关于层次划分（致编辑的信）》，美国统计学家46(1)，70–71。

Tetteh, J., Ekem-Ferguson, G., Quarshie, E. N., Swaray, S. M., Ayanore, M. A., Seneadza, N. A. H., Asante, K. O. 和 Yawson, A. E.（2021），《非洲在校青少年大麻使用与自杀行为：基于全球学校学生健康调查对患病率和风险因素的评估》，综合精神病学 34(4)。

Vos, P. W.（1991），‘一种检测影响性案例的几何方法’，统计年刊19(3)，1570–1581。

Walsh, C. J., Papas, P. J., Crowther, D., Sim, P. T. 和 Yoo, J.（2004），‘雨水排水管对澳大利亚东南部具有保护意义的溪流栖息端足类*Austrogammarus australis*构成的威胁’，《生物多样性与保护》13(4)，781–793。

Yip, K. C. H. 和 Yau, K. K. W.（2005），‘在一般保险中对含额外零值的索赔频率数据建模’，保险：数学与经济学36(2)，153–163。

Yu, T., Zhou, H., Suh, K. 和 Arcona, S.（2015），‘评估慢性阻塞性肺疾病患者非计划性医院再入院中预测变量的重要性’，*ClinicoEco-nomics and Outcomes Research*7，37–51。

Zhang, Y.（2013年），《Tweedie复合泊松线性混合模型的似然和贝叶斯方法》，统计与计算**23**，743–757。

拟合优度度量与（伪）Shapley 值之间的关系

本附录概述了四种拟合优度度量 R^2_{KL}、 R^2_{McF}、$\ell - \ell(\varnothing)$ 和 ℓ 与相应（伪）Shapley 值之间的关系。

首先，由于这些量之间均呈线性关系，我们可以一并研究它们之间的关系。具体而言，考虑一组回归变量 P 和两个呈线性关系的拟合优度度量 $v_2(S) = a\, v_1(S) + b$，其中 a、 $b \in \mathbb{R}$ 和 $S \subseteq P$ 是可用回归变量的某个子集。根据方程（1），偏移项 b 会消去，因此得到的（伪）Shapley 值彼此成比例，即 $\varphi_i(P,\ v_2) = a\,\varphi_i(P,\ v_1)$。另请注意，如果拟合度量 v 不满足 $v(\varnothing) = 0$，则会得到伪 Shapley 值 $\varphi^*_i(P,\ v)$。

由方程（13）和（11）可得，给定数据 $\mathbf{y}$ 以及一组子集 S 回归变量所产生的预测值 $\hat{}\mu_S$，

$$R^2_{\mathrm{KL}}(\hat{\mu}_S) \;=\; \frac{\ell(\mathbf{y},\hat{\mu}_S) - \ell(\mathbf{y},\hat{\mu}_0)}{\ell(\mathbf{y},\mathbf{y}) - \ell(\mathbf{y},\hat{\mu}_0)}, \tag{30}$$

其中 $\ell(\mathbf{y},\mathbf{y})$ 表示饱和模型的对数似然。与前面一样，将这一恒等式用相关子集 S表示更为方便，使用 $R^2_{\mathrm{KL}}(S) = R^2_{\mathrm{KL}}(\hat{}\mu_S)$和 $\ell(S) = \ell(\mathbf{y},\hat{}\mu_S)$，其中 $S \subseteq P$。如第 3.3 节所示，令 P'表示饱和模型。于是，方程（30）可改写为

$$R^2_{\mathrm{KL}}(S) \;=\; \frac{1}{\ell(P') - \ell(\varnothing)}\,\ell(S) - \frac{\ell(\varnothing)}{\ell(P') - \ell(\varnothing)}; \tag{31}$$

由此可得

$$\varphi_i(P, R^2_{\mathrm{KL}}) \;=\; \frac{1}{\ell(P') - \ell(\varnothing)}\,\varphi^*_i(P,\ell). \tag{32}$$

Sim类似地，将方程（17）改写为回归变量集合的形式，可得 o

$$R^2_{\mathrm{KL}}(S) \;=\; \frac{-\ell(\varnothing)}{\ell(P') - \ell(\varnothing)}\,R^2_{\mathrm{McF}}(S); \tag{33}$$

由此可得

$$\varphi_i(P, R^2_{\mathrm{KL}}) \;=\; \frac{-\ell(\varnothing)}{\ell(P') - \ell(\varnothing)}\,\varphi_i(P, R^2_{\mathrm{McF}}). \tag{34}$$

最后， $\ell(S) - \ell(\varnothing)$ 与 $\ell(S)$ 之间的关系显然是线性的，其中 $a = 1$，因此

$$\varphi_i(P, \ell - \ell(\varnothing)) \;=\; \varphi^*_i(P,\ell). \tag{35}$$

请注意，使用 $R^2_{\mathrm{KL}}(S)$、$R^2_{\mathrm{McF}}(S)$ 或 $\ell - \ell(\varnothing)$ 会得到 Shapley 值，而使用 $\ell(S)$ 则会得到伪 Shapley 值。

拟合优度度量会随着新增回归变量而单调递增。因此，方程（31）中的乘法常数为正，从而建立了与正Shapley值之间的映射。方程（33）中乘法常数的符号则不那么显而易见，因为对数似然的符号取决于具体的模型和数据。然而，对于二元响应模型，对数似然始终非正；因此，这些模型中方程（33）里的乘法常数非负。（事实上，它等于一，因为在这种情况下，McFadden似然比指数 R^2_{McF} 等于 R^2_{KL} 。）