arXiv:2611.32792v1 [cs.AI] 24 Nov 2026

脑分数噪声上限估计量的准确性

Aude Maier Lucas Gruaz MartinSchrimpf* Johanni Brea*

洛桑联邦理工学院，洛桑，瑞士

{aude.maier,lucas.gruaz,martin.schrimpf,johanni.brea}@洛桑联邦理工学院.ch

摘要

现代机器学习模型的内部激活与真实脑中记录的活动具有出人意料的良好一致性。在过去五年中，编码模型与脑活动之间的这种一致性（通常称为脑分数）稳步提升。由于脑记录存在噪声，任何模型所能达到的脑分数都存在一个上界。这个上界被称为噪声上限，具有两方面的关键作用：它决定了模型性能是否仍有进一步提升的可能，也使跨基准测试比较脑分数变得有意义。该领域已经发展出多种估计噪声上限的方法，但现有估计量的统计性质仍缺乏充分理解，因此尚不清楚它们在哪些实验设置下能够提供可靠的估计。在本文中，我们为脑记录引入了一个统一的数学框架，利用该框架推导出四种常用噪声上限估计量的闭式表达式，并对每种估计量的行为给出理论预测。我们通过模拟验证这些理论预测，并将模拟结果与现有功能性磁共振成像和脑电图数据集中的噪声上限估计进行比较。我们发现，在不同设置下，现有方法都会低估真实噪声上限。最后，我们针对不同情况提出最佳实践建议。

1 引言

神经人工智能的一个核心目标，是构建不仅能够解决大脑所解决的任务，而且其内部表征也与生物神经回路的表征相匹配的计算模型（Yamins and DiCarlo, 2016；Schrimpf 等, 2020；Doerig 等, 2023；Sadeh and Clopath, 2025）。用于检验这种匹配关系的主流实证工具是脑分数：它衡量模型的内部激活在多大程度上能够预测神经反应，所依据的是同一刺激集，通常通过在训练集划分上拟合线性读出，并在留出数据上进行评估（例如 Schrimpf 等, 2018）。自首次证明经过任务优化的卷积神经网络比手工设计模型更能预测腹侧通路反应以来，脑分数已在多种模态中广泛应用：视觉（Yamins 等, 2014；Khaligh-Razavi以及Kriegeskorte，2014；Schrimpf 等，2018；Cadena 等，2019）、听觉（Kell 等，2018；Tuckute 等，2023）和语言（Wehbe 等，2019；Schrimpf 等，2021；Caucheteux 等，2023；Lei 等，2026；AlKhamissi 等，2025）。随着架构和训练数据规模的扩大，它们取得了显著改进（Antonello 等，2023；Gokce 和 Schrimpf，2025）。

脑评分应相对于可达到的水平进行衡量。神经记录存在噪声：试次间变异性、测量噪声和有限采样，都对任何确定性模型所能预测的方差设置了严格上限（Hsu 等人，2004；Sahani 和 Linden，2002；Schoppe 等人，2016）。噪声上限正是区分一个已捕捉到刺激驱动信号的模型与一个仍有改进空间的模型的依据。没有噪声上限，绝对脑评分很难解释，因为它们将模型质量与数据可靠性混为一谈。噪声上限将问题从“分数有多高

*共同指导

第40届神经信息处理系统大会（NeurIPS 2026）。

？”重新表述为“可解释方差中有多少已经得到解释？”，这是一个可以在不同数据集、区域和记录模态之间公平比较模型的量。

然而，估计噪声上限比看起来更加困难。相关文献包含多种方法。Sahani 和 Linden（2002）提出了可解释信号功率估计量（SPE），用于衡量模型所捕获的神经反应中刺激相关方差的比例。Hsu 等人（2004）提出了归一化相关系数 CC_{norm}，该系数将模型性能表示为所估计上界 CC_{max} 的比例；这一上界通过采用 Spearman–Brown 校正的分半信度计算得到。Schoppe 等人（2016）表明，在某些设定下，CC_{norm} 优于 SPE，并推导出一种无需重采样、直接根据神经信号功率计算 CC_{max} 的替代方法，从而避免分半程序的计算开销。Lage-Castellanos 等（2019）将这一分析扩展至功能性磁共振成像，比较了不同噪声结构和正则化设定下的蒙特卡罗、分半法和解析噪声上限估计量。Schrimpf 等（2021）提出了一种用于 Brain-Score 的外推方法。van Bree 等（2025）澄清了模型评估中对噪声上限的一种常见误解，指出分半信度常被错误地视为可达到的最大相关性，而非可解释方差的上界（R^2）。他们证明，这种混淆可能导致对模型性能的得分产生虚高或误导性结果。

在这项工作中，我们从根本上重新审视噪声上限。我们回顾估计量的主要类别，在脑信号生成模型下推导解析表达式，并刻画每种估计量的假设与失效模式。通过实证分析，我们展示了在功能性磁共振成像和脑电图数据上，模型预测与噪声上限估计之间的定性一致性。最后，我们就应使用哪种估计量提出具体建议，旨在为 brain-score 比较奠定更坚实的统计基础。

2 背景

从刺激驱动实验中获得的脑记录可以表示为一组向量$\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}, \mathbf{v}_i \in \mathbb{R}^T$。每个向量 $\mathbf{v}_i$ 对应于单个记录单元的反应，例如，功能性磁共振成像中的一个体素或脑电图中的一个电极。如果实验涉及单个连续刺激（例如观看电影或聆听故事），T 表示刺激持续时间除以记录设备的时间分辨率，而$v_{i,t}$ 表示记录单元 i 在时间步 t上的信号。对于由一系列离散刺激组成的实验，每个索引 t 代表一个刺激，而 T 表示刺激总数。噪声上限的计算依赖于这样的实验：相同的刺激会被呈现多次，并利用由此产生的试次来评估脑反应的一致性。我们用 $\mathbf{v}_i^{(\ell)}, \ell \in \{1, \ldots, n\}$ 表示记录单元 i 在试次 ℓ期间记录的信号。根据实验设置的不同，试次可以指向同一被试对相同刺激的多次呈现，也可以指向向不同被试各呈现一次刺激。尽管这两种设置具有不同的统计解释（我们将在后文讨论），但它们可以在统一框架下进行处理。

编码模型与脑评分 编码模型旨在建立从刺激空间到脑记录空间的映射，即根据刺激序列 $t = 1, \ldots, T$，预测每个记录单元 i的脑反应 $\hat{\mathbf{v}}_i$。在本研究中，我们聚焦于编码模型，不考虑表征相似性分析（RSA）框架。此类模型的性能通常通过两步程序进行评估：i）计算每个单元 i 的预测值 $\hat{\mathbf{v}}_i$ 与实际记录数据 $\mathbf{v}_i$之间的相似性 S_i；ii）计算平均得分 $S = \frac{1}{m}\sum_{i=1}^{m} \frac{S_i}{\text{NC}_i}$，并将其与单元 i的噪声上限估计值 NC_i 进行比较。最常用的相似性度量是皮尔逊相关[2] $S_i = \widehat{\text{Corr}}_t(\hat{\mathbf{v}}_i, \mathbf{v}_i)$。相关性可以在单个试次层面计算，$\widehat{\text{Corr}}_t\left(\hat{\mathbf{v}}_i^{(\ell)}, \mathbf{v}_i^{(\ell)}\right)$，也可以基于跨试次平均的信号计算 $\widehat{\text{Corr}}_t(\hat{\mathbf{v}}_i, \bar{\mathbf{v}}_i)$。这两种方法的选择通常取决于实验设置：被试内重复通常会取平均以减少噪声，而跨被试试次则更常保持分开处理，既是为了保留个体特异性信号，也是因为难以在不同被试之间建立记录单元的有意义对应关系。编码模型可以通过在记录信号的留出部分上进行训练，为每个被试学习一个单独的映射。然而，如果记录数据包含既非由刺激确定性驱动、也未被试次特定参数捕获的变异性，则任何编码模型都无法达到完美得分。

[2] $\widehat{\text{相关系数}}_t(\mathbf{x}, \mathbf{y}) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \mu_x)(y_t - \mu_y)/(\sigma_x \sigma_y)$，随时间计算的平均值为 μ_x，经验方差为 σ_x^2

噪声上限 如果我们能够获得无噪信号 $\mathbf{v}_i^{*(\ell)}$，就可以计算记录单元 i 的噪声上限，计算如下：

$$NC_i^{\text{mean}} = \widehat{\text{Corr}}_t\left(\overline{\mathbf{v}}_i^*, \overline{\mathbf{v}}_i\right), \qquad NC_{i,\ell}^{\text{single}} = \widehat{\text{Corr}}_t\left(\mathbf{v}_i^{*(\ell)}, \mathbf{v}_i^{(\ell)}\right), \tag{1}$$

具体取决于编码模型预测的是试次平均响应还是单个试次。对于给定数据集 $\{\mathbf{v}_1^{(\ell)}, \ldots, \mathbf{v}_m^{(\ell)}\}_{\ell=1}^n$，该噪声上限将构成任何编码模型所能达到的得分上界。然而，由于无噪声信号 $\mathbf{v}_i^{*(\ell)}$ 未知，因此无法直接计算噪声上限。已有多种方法被提出，用于仅根据观测记录来估计噪声上限。文献中通常使用其中五种方法，这些方法可分为两类：一类以 NC^{mean} 为目标，另一类以 NC^{single} 为目标。表 3 列出了使用这些方法的论文。我们将在本节余下部分分别介绍这些方法。

2.1 试次平均响应的噪声上限

通常使用两种方法来估计 NC^{mean}。这两种方法都假设每条记录可分解为 $\mathbf{v}_i^{(\ell)} = \mathbf{v}_i^* + \boldsymbol{\eta}_i^{(\ell)}$，其中 $\mathbf{v}_i^*$ 在各个试次之间完全共享，而 $\boldsymbol{\eta}_i^{(\ell)}$ 是独立同分布的，其方差为 σ_η^2（Hsu 等人，2004；Schoppe 等人，2016）。在该模型下，当记录足够长时，真实噪声上限为

$$NC_i^{\text{mean}} = \widehat{\text{Corr}}_t\left(\mathbf{v}_i^*, \mathbf{v}_i^* + \overline{\boldsymbol{\eta}}_i\right) = \sqrt{\frac{\sigma_s^2}{\sigma_s^2 + \frac{\sigma_\eta^2}{n}}}, \tag{2}$$

其中 $\sigma_s^2 = \widehat{\text{Var}}_t(\mathbf{v}_i^*)$。在 **斯皮尔曼-布朗校正的半分信度（SHSB）**方法中，将 n 试次划分为两个相等的部分，并对各自的平均值进行相关分析。由于每个半部分的平均值所包含的噪声是完整平均值的两倍，因此需要应用斯皮尔曼-布朗校正（Hsu 等人，2004；van Bree 等，2025）。该过程在多次随机划分上重复进行，并对结果取平均值（表 1）。另一种方法（Schoppe 等人，2016；Lage-Castellanos 等，2019；Allen 等，2022）——**跨试次方差方法（ATV）**——使用 $\sigma_s^2 = \sigma_{\text{total}}^2 - \sigma_\eta^2/n$ 估计信号方差，其中总方差 σ_{total}^2 估计为 $\widehat{\text{Var}}_t(\mathbf{v}_i)$，噪声方差 σ_η^2 则取每个时间步的跨试次方差平均值（表 1）。

2.2 单试次的噪声上限

一些实验对每位受试者仅进行一次较长的连续记录（例如电影或故事（Freteault 等，2025 年；Chen 等，2017 年；Wehbe 等，2019 年）），其中试次对应于不同的受试者。对于低变异性的感觉区域，研究人员会对受试者进行平均，并应用 SHSB 或 ATV（Tang 等，2025 年；Tuckute 等，2024 年）。对于高级区域（例如语言区域），被试间变异性很大，解剖错位可能会抵消共享信号，因此需要针对每位受试者拟合编码模型，并改为估计 NC^{single}。我们总结三种常用的估计量。**留一相关（LOOC）**：将每个试次 $\mathbf{v}_i^{(\ell)}$ 与留一平均值 $\mathbf{v}_i^{(-\ell)} := \frac{1}{n-1}\sum_{\ell' \neq \ell} \mathbf{v}^{(\ell')}{}_i$ 进行相关分析，并对所有试次的结果取平均。**试次对均值相关（TTMC）**：与 LOOC 相同，但个体试次也会被纳入与其进行比较的平均值中。**留一预测（LOOP）**由 Schrimpf 等（2021）提出：在线性模型中，使用由 n 试次组成的数据池来预测留出试次 $\mathbf{v}_i^{(\ell)}$，从而模拟编码模型的训练过程。数据池大小 n 会发生变化，并通过拟合 $\widehat{NC}_{i\,\ell}^{\text{LOOP}}{}_{,}(n) = \widehat{NC}_i^{\infty}{}_{,\ell}\left(1 - e^{-n/\tau_0}\right)$ 将噪声上限外推至 $n \to \infty$（表 1）。

3 结果

3.1 脑记录数据的生成模型

我们将一次记录 $\mathbf{v}_i^{(\ell)} = \{v_{i,t}^{(\ell)}\}_{t>0}$ 定义为一个随机过程。每个试次 ℓ 都具有属性 $\theta^{(\ell)}$，这些属性在试次内固定，但在试次之间可能不同，例如实验设置、受试者的内部状态或受试者的身份。我们将信号的可重复部分定义为条件期望 $v_i^{*(\ell)}{}_{,\,t} = \mathbb{E}\left[v_{i,t} \mid \theta^{(\ell)}\right]$ 该期望是对试次内不具可重复性的一切因素求取的，而将噪声定义为 $\eta_i^{(\ell)}{}_{,\,t} = v_i^{(\ell)}{}_{,\,t} - v_i^{*(\ell)}{}_{t\,,}$。属性 $\theta^{(\ell)}$ 仅当试次是在相同条件下重复呈现时，才会对所有 ℓ 保持相同；在这种情况下，$v_{i,\,t}^*$ 对每个试次都相同。

表 1：噪声上限估计量公式。平方根下的负值将按零处理；SHSB 还会对多个随机划分取平均值。$\mathbf{v}$：试次平均值，$\mathbf{v}_i^{(-\ell)}$：排除试次 ℓ后的试次平均值，$\mathbf{v}_i^{(1)}$：前一半（随机选取）试次的平均值，$\widehat{\mathbf{v}}_i^{(\ell)}$：给定所有其他试次时的线性预测

	试次平均 $\left(\widehat{NC}_i^{\text{mean}}\right)$		单试次 $\left(\widehat{NC}_{i,\ell}^{\text{single}}\right)$
SHSB	$\sqrt{\dfrac{2\,\widehat{\text{相关性}}_t\left(\mathbf{v}_i^{(1)},\,\mathbf{v}_i^{(2)}\right)}{1+\widehat{\text{相关性}}_t\left(\mathbf{v}_i^{(1)},\,\mathbf{v}_i^{(2)}\right)}}$	LOOC	$\widehat{\text{相关性}}_t\left(\mathbf{v}_i^{(\ell)},\,\mathbf{v}_i^{(-\ell)}\right)$
		TTMC	$\widehat{\text{相关性}}_t\left(\mathbf{v}_i^{(\ell)},\,\mathbf{v}_i\right)$
ATV	$\sqrt{\dfrac{\widehat{\text{方差}}_t(\mathbf{v}_i)-\frac{1}{n}\frac{1}{T}\sum_t\widehat{\text{方差}}_\ell\left(v_{i,t}^{(\ell)}\right)}{\widehat{\text{Var}}_t(\mathbf{v}_i)}}$	LOOP	$\widehat{\text{相关性}}_t\left(\mathbf{v}_i^{(\ell)},\,\widehat{\mathbf{v}}_i^{(\ell)}\right)$

中心化可重复响应 $\mathbf{v}_i^{*(\ell)}-\langle\mathbf{v}_i^{*(\ell)}\rangle$（其中 $\langle\cdot\rangle$ 表示对时间求平均值）张成了 $\mathbb{R}^T$ 的一个子空间，其维度为 $K\le nm$。对于该子空间的任意一组基 $\{\boldsymbol{\phi}_k\}_{k=1}^K$，若其在时间上正交归一，即当 $\langle\boldsymbol{\phi}_k\boldsymbol{\phi}_{k'}\rangle=1$时为 $k=k'$，否则为 0，则单元 i 在试次 ℓ期间记录到的信号可写为

$$\mathbf{v}_i^{(\ell)}=b_i^{(\ell)}+\sum_{k=1}^{K}w_{ik}^{(\ell)}\boldsymbol{\phi}_k+\boldsymbol{\eta}_i^{(\ell)},\qquad \mathbf{v}_i^{*(\ell)}=b_i^{(\ell)}+\sum_{k=1}^{K}w_{ik}^{(\ell)}\boldsymbol{\phi}_k, \tag{3}$$

其中 $b_i^{(\ell)}:=\langle\mathbf{v}_i^{*(\ell)}\rangle$ 是一个常数偏移，$w_{ik}^{(\ell)}:=\widehat{\text{协方差}}_t(\mathbf{v}_i^{*(\ell)},\,\boldsymbol{\phi}_k)$ 是试次和单元特异性的混合权重，$\boldsymbol{\eta}_i^{(\ell)}$ 是上文定义的噪声。

潜在信号 这种线性分解提供了一种有用的解释：基信号 $\boldsymbol{\phi}_k$可以看作由刺激诱发的神经活动的潜在信号。这一解释建立在这样的观点之上：刺激驱动的脑活动存在于完整神经活动空间的低维子空间中（Cunningham 和 Yu, 2014；Gao 等, 2017），因此，K 通常远小于神经元的数量，尽管它可能超过所测量单元的数量 m。此外，跨被试相关文献表明，刺激驱动的响应在很大程度上由各被试共享（Hasson 等, 2004；Simony 等, 2016），这意味着存在一个所有被试共有的神经活动子空间。这种表述类似于共享响应模型（Chen 等, 2015）及相关的功能对齐方法（Haxby 等, 2020）；这些方法将功能性磁共振成像记录中的刺激驱动成分分解为共享的潜在信号和被试特异性的混合权重。这也是编码模型所预设的内容：单个模型生成各被试共享的特征，并为每个被试拟合一个被试特异性的线性读出，这意味着可重复响应位于由模型特征张成的共同子空间中。在这一视角下，K 是模型所需特征数量的下界。该框架有意设计为适用于不同的记录模态。

混合权重和状态 权重 $w_{ik}^{(\ell)}$ 描述了每个潜在信号 $\boldsymbol{\phi}_k$在试次 ℓ期间单元 i 的可重复响应中表达的强度。在理想情况下，每个试次都在完全相同的条件下（即属性 $\theta^{(\ell)}$ 为常数）向同一受试者呈现相同的刺激，权重在每个试次 $w_{ik}^{(\ell)}=w_{ik}$中都相同。我们将其称为同质状态。当试次对应于不同被试时，解剖结构和功能组织的差异会导致每个被试以不同方式表达潜在信号。即使对于同一被试，实验设置或被试内部状态的变化，也会改变潜在信号在不同试次中的表达方式。在这两种情况下，权重都会因试次而异，我们称之为异质状态。因此，应将同质状态视为一种理想化情形。对同一被试进行的重复试次可以接近这一状态，而从不同被试获取的试次通常不满足这一条件。

噪声 噪声项 $\boldsymbol{\eta}_i^{(\ell)}$ 涵盖记录信号中所有在试次内无法复现的变异性。它可能源于多个方面，例如神经过程的内在随机性、

不受刺激驱动的持续活动，或记录设备引入的噪声。它可能表现出时间相关性，其方差可能取决于信号，并随时间推移或在不同试次之间发生变化。

将信号模型和式（3）中的无噪声分量代入第 2 节所介绍的噪声上限估计量后，我们可以推导出它们的闭式表达式。这使我们能够针对每种方法评估其在何种条件下可以准确估计真实噪声上限，以及在哪些情况下可能存在不足。我们将在下文中开展这一分析，依次讨论 NC^{mean}和 NC^{single}估计量。

符号与假设 对于每个单元 i，我们记 $\bar{w}_{ik} := \frac{1}{n}\sum_{\ell=1}^{n} w_{ik}^{(\ell)}$ 为跨试次的平均权重，记 $\widehat{\mathrm{Var}}_{\ell}\left[w_{ik}\right] := \frac{1}{n}\sum_{\ell=1}^{n}\left(w_{ik}^{(\ell)} - \bar{w}_{ik}\right)^2$ 为其方差。类似地，我们记 $\sigma_{i,\ell}^2 := \frac{1}{T}\sum_{t=1}^{T}\mathrm{Var}\left(\eta_{i,t}^{(\ell)}\right)$ 为试次 ℓ 在时间上平均后的噪声方差，并记$\bar{\sigma}_i^2 := \frac{1}{n}\sum_{\ell=1}^{n}\sigma_{i,\ell}^2$ 为其跨试次平均值。所有解析结果都依赖于附录 B.2 中形式化陈述的四个假设：（1）不同试次的噪声 $\boldsymbol{\eta}_i^{(\ell)}, \boldsymbol{\eta}_i^{(\ell')}$ 在给定属性 $\{\theta^{(\ell)}\}_{\ell=1}^{n}$ 的条件下相互独立（同一试次内不同单元的噪声可能相关）；（2）无噪声信号 $v_{i\,t}^{*(\ell)}$ 有界；（3）噪声 $\eta_{i\,t}^{(\ell)}$ 具有绝对可和的时间自协方差，即其随时间变化的自相关快速衰减；（4）噪声平方 $\eta_{i}^{(\ell)}{}_{,t}{}^{2}$ 具有有限方差和绝对可和的时间自协方差。第 4 节将讨论这些假设。此外，下面的闭式表达式仅适用于刺激驱动部分非零的单元 i，即对于试次平均噪声上限，$\widehat{\mathrm{Var}}_t\left(\bar{\mathbf{v}}_i^{*}\right) > 0$，对于单试次噪声上限，$\widehat{\mathrm{Var}}_t\left(\mathbf{v}_i^{*(\ell)}\right) > 0$，从而式（1）中的各个量都有明确定义。

3.2 试次平均响应的噪声上限

如附录B.4.1所推导，单元 i 的试次平均响应的真实噪声上限为

$$NC_i^{\mathrm{mean}} = \sqrt{\frac{\sum_{k=1}^{K}\bar{w}_{ik}^2}{\sum_{k=1}^{K}\bar{w}_{ik}^2 + \bar{\sigma}_i^2/n}} + \varepsilon_T, \tag{4}$$

其中，ε_T是一个具有偏差 $|\mathbb{E}\left[\varepsilon_T\right]| \le c/T$和均方误差 $\mathbb{E}\left[\varepsilon_T^2\right] \le c/T$的随机误差项，$c$ 为仅取决于我们假设中的界的常数。对于较长的记录数据，NC_i^{mean}趋近于公式（4）中的主导项，其围绕该项的随机波动以 $T^{-1/2}$的速率缩小，系统偏差则以 T^{-1}的速率缩小。可以观察到，当 $n \to \infty$ 时，噪声上限收敛于1。注意，在同质状态下，$\widehat{\mathrm{Var}}_{\ell}[w_{ik}] = 0$，且公式（4）收敛于公式（2）。

ATV 估计量 如附录B.4.2所推导，ATV 估计量为

$$\widehat{NC}_i^{\mathrm{ATV}} = \sqrt{\left[\frac{\sum_k \bar{w}_{ik}^2 - \frac{1}{n-1}\sum_k \widehat{\mathrm{Var}}_{\ell}\left[w_{ik}\right]}{\sum_k \bar{w}_{ik}^2 + \bar{\sigma}^2/n}\right]_+} + \varepsilon_T^{\max} = \sqrt{\left[\left(NC_i^{\mathrm{mean}}\right)^2 - \frac{\frac{1}{n-1}\sum_k \widehat{\mathrm{Var}}_{\ell}\left[w_{ik}\right]}{\sum_k \bar{w}_{ik}^2 + \bar{\sigma}^2/n}\right]_+} + \varepsilon_T^{\max}, \tag{5}$$

其中，$[x]_+ := \max(0, x)$，且随机误差项满足 $|\mathbb{E}\left[\varepsilon_T^{\max}\right]| \le (c/T)^{1/4}$ 以及$\mathbb{E}\left[\left(\varepsilon_T^{\max}\right)^2\right] \le (c/T)^{1/2}$，其中 c为既不取决于 T也不取决于 n的常数。这些界弱于 ε_T的界，因为阈值 $\max(0, \cdot)$在零点不可微。设平方根的自变量为 x_i，如果 x_i 关于 T一致地有界离零点有一定距离，则误差项变为 ε_T，并恢复 $1/T$ 速率。然而，只有当 x_i 估计的波动相对于 x_i足够小时，才能达到这一速率。对于信号较弱的单元（ x_i较小），在 T 较小时，这些波动相对于 x_i 与零之间的距离较大，误差为 $T^{-1/4}$量级。在这一瞬态区域中，误差主要由向上偏差主导，因为阈值会将随机偏低的实现值设为零，却保留随机偏高的实现值。此外，$\widehat{NC}_i^{\mathrm{ATV}}$ 的主导项会以 $\mathcal{O}(1/n)$量级低估真实噪声上限，而该项在同质状态下消失。

SHSB估计量 我们将 $(A,\ B)$ 记作将 n 试次划分为两个大小为 $n/2$ 的半部分，并将 $\mathcal{S}$ 记作对其进行斯皮尔曼–布朗校正相关平均的 r 划分集合。如

附录 B.4.3 中所推导的，如果对于每个划分 $\widehat{\mathrm{Cov}}_t\left(\mathbf{v}_i^{*A},\ \mathbf{v}_i^{*B}\right) \geq 0$，那么

$$\widehat{NC}_i^{\mathrm{SHSB}} = \sqrt{\frac{1}{r}\sum_{(A,B)\in\mathcal{S}} \frac{\sum_k \bar{w}_{ik}^2 - \Delta_{AB}^{w}}{\sum_k \bar{w}_{ik}^2 + \bar{\sigma}_i^2/n - \Delta_{AB}^{\mathrm{sd}}}} + \varepsilon_T^{\max}, \tag{6}$$

其中 $\Delta_{AB}^{w} := \frac{1}{4}\sum_k \left(\bar{w}_{ik}^A - \bar{w}_{ik}^B\right)^2$ 且 $\Delta_{AB}^{\mathrm{sd}} := \frac{1}{4}\left(\sqrt{\sum_k \left(\bar{w}_{ik}^A\right)^2 + \frac{2\bar{\sigma}_A^2}{n}} - \sqrt{\sum_k \left(\bar{w}_{ik}^B\right)^2 + \frac{2\bar{\sigma}_B^2}{n}}\right)^2$，

其中 $\bar{\sigma}_A^2 := \frac{2}{n}\sum_{\ell\in A}\sigma_{i,\ell}^2$（$B$也类似）。项 Δ_{AB}^{w} 衡量两半之间潜在权重的差异程度，而 $\Delta_{AB}^{\mathrm{sd}}$ 衡量两个半平均值随时间变化的标准差之间的差异程度。误差项 $\varepsilon_T^{\max}$的表现与 ATV 估计量相同：当平方根内的参数在 T上均匀地与零保持有界距离时，它为 ε_T；但对于信号较弱的单元，在较小的 T下，其量级为 $T^{-1/4}$。

SHSB 和 ATV 估计量密切相关。当 $\mathcal{S}$ 包含所有可能的划分时，Δ_{AB}^{w} 在各划分上的平均值恰好为 $\frac{1}{n-1}\sum_k \widehat{\mathrm{Var}}_\ell\left[w_{ik}\right]$（引理5）。因此，平均后的分子与公式（5）的分子一致，并且两个估计量都具有相同的、由跨试次的信号异质性导致的 $\mathcal{O}(1/n)$ 低估。它们的差异仅体现在分母中的 $\Delta_{AB}^{\mathrm{sd}}$。

3.3 单试次反应的噪声上限

附录 B.5.1 中推导出的真实噪声上限 $NC_{i,\ell}^{\mathrm{single}}$，给出如下：

$$NC_{i,\ell}^{\mathrm{single}} = \sqrt{\frac{\sum_k \left(w_{ik}^{(\ell)}\right)^2}{\sum_k \left(w_{ik}^{(\ell)}\right)^2 + \sigma_\ell^2}} + \varepsilon_T. \tag{7}$$

与试次平均情形不同，它不依赖于试次次数 n，而仅由单个记录单元的信噪比决定。

LOOC 和 TTMC 估计量 附录 B.5.2 和 B.5.3 中详细介绍的 LOOC 和 TTMC 估计量表达式为

$$\widehat{NC}_{i,\ell}^{\mathrm{LOOC}} = NC_{i,\ell}^{\mathrm{single}} \cdot \frac{\sum_k w_{ik}^{(\ell)}\bar{w}_{ik}^{(-\ell)}}{\sqrt{\left(\sum_k \left(w_{ik}^{(\ell)}\right)^2\right)\left(\sum_k \left(\bar{w}_{ik}^{(-\ell)}\right)^2 + \frac{\bar{\sigma}_{(-\ell)}^2}{n-1}\right)}} + \varepsilon_T,$$

$$\widehat{NC}_{i,\ell}^{\mathrm{TTMC}} = NC_{i,\ell}^{\mathrm{single}} \cdot \frac{\sum_k w_{ik}^{(\ell)}\bar{w}_{ik} + \frac{\sigma_\ell^2}{n}}{\sqrt{\left(\sum_k \left(w_{ik}^{(\ell)}\right)^2\right)\left(\sum_k \bar{w}_{ik}^2 + \frac{\bar{\sigma}^2}{n}\right)}} + \varepsilon_T. \tag{8}$$

当 n变大时，两个表达式都趋近于 $NC_{i,\ell}^{\mathrm{single}} \cdot \cos_k\left(w_i^{(\ell)}, w_i\right) + \varepsilon_T$，其中 $\cos_k(\cdot,\ \cdot)$ 表示跨 k的余弦相似度。然而，在异质设置下，该极限严格低于 NC_i^{single}。值得注意的是，这种偏差源于编码模型会为每个试次生成一个单独的预测。其他在所有试次间使用同一个共享预测的方法，例如表征相似性分析（RSA）（Kriegeskorte 等，2008），则不受此偏差影响。在同质状态下，随着 $n \to \infty$，两个估计量都收敛于真实噪声上限 $NC_{i,\ell}^{\mathrm{single}}$，直至 ε_T。LOOC 估计量始终从下方趋近于该值。只要试次 ℓ 的噪声方差至少为试次间平均噪声方差的一半，$\sigma_\ell^2 \geq \bar{\sigma}^2/2$TTMC 估计量就会从上方趋近于该值。

LOOP估计量 LOOP无法用权重表示为闭式形式，但其行为可以从几何角度进行描述（附录 B.5.4）。当 T 相对于预测变量的数量 $p = m\,(n-1)$ 较大时，$\widehat{NC}_{i,\ell}^{\mathrm{LOOP}} = NC_{i,\ell}^{\mathrm{single}} \cos\,\alpha^{(\ell)}$，其中 $\alpha^{(\ell)}$ 是 $\mathbf{v}_i^{*(\ell)}$ 与预测变量张成的空间之间的夹角。因此，LOOP 是单试次噪声上限的下界；至少在 T/p 较大的极限下，它不小于 LOOC。随着 n 增大且保持 T/p 较大，$\alpha^{(\ell)}$ 减小，LOOP 最终会达到真实噪声上限，即使在异质状态下也是如此。在实践中，

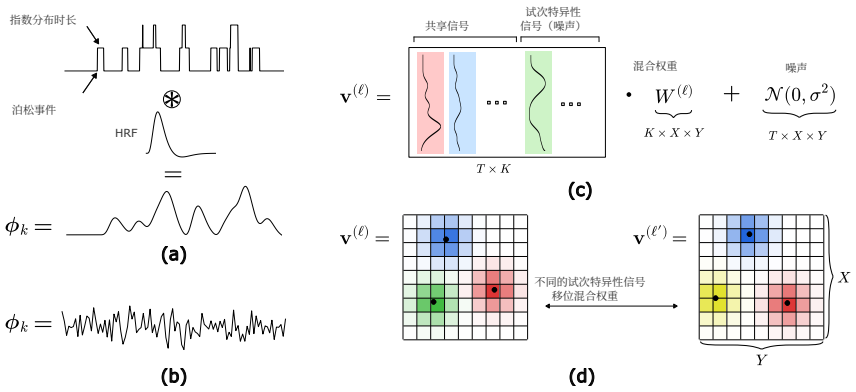

图 1：**信号与试次生成**。（a）为模拟连续刺激引起的功能磁共振成像血氧水平依赖响应，通过将事件时间序列与血流动力学响应函数（HRF）进行卷积来生成信号。（b）为模拟一组随机排序的离散刺激引起的响应，将信号生成为独立同分布的高斯过程。（c）一个试次通过将共享信号和试次特异性信号与试次特异性混合权重相乘，并加入高斯噪声而生成。（d）两个生成的试次因其试次特异性信号以及（可能）未对齐的混合权重而有所不同。关于生成过程的更多细节见第 3.4 节。

然而，时间点的数量 T 受到限制，并且拟合模型的泛化误差会随 p 增大，因此 LOOP 在实践中可能低于 LOOC。因此，它对 n 的依赖结合了一个递增项和一个递减项，不存在简单的通用形式。特别是，指数形式 $1 - e^{-n/\tau_0}$ 假设会快速达到饱和，并忽略了泛化误差；这两点都会导致低估噪声上限，也可能解释了为什么采用该方法估计的噪声上限常常低于所报告的编码模型得分（Aw 等，2024）。

3.4 实验结果

本节将在合成脑记录以及真实功能性磁共振成像和脑电图数据上评估噪声上限估计量。合成数据集作为通用数学框架的一种特定设置构建而成，可对真实噪声上限、第 3.2 节和第 3.3 节推导出的闭式表达式，以及经验估计进行定量比较。真实数据集则在不同模态和实验设置下，对理论结果和合成结果进行定性验证。除 LOOP 外，所有结果均在单个 CPU 上于一天内完成计算。LOOP 则在单个 NVIDIA A100 GPU（80 GB）上耗时约一天完成计算。用于生成所有图表的代码可在 https://github.com/AudeMaier/ noise-ceiling 获取。

功能性磁共振成像和脑电图数据集 我们在三个数据集上评估噪声上限估计量。在每种情况下，我们分别从 n 个试次、 m 个记录单元以及长度为 T的刺激序列这几个方面描述数据。

- **THINGS fMRI1：** 一个功能性磁共振成像数据集（Hebart 等，2023），其中3名参与者在 3T 磁共振成像扫描仪上接受扫描，同时观看来自 THINGS 数据库（Hebart 等，2019）的图像序列。每名参与者进行了12次会话，在此期间重复呈现了100张图像。我们使用一般线性模型（GLM）对第一名参与者V1区体素获得的响应估计值（β 值）。所得数据集包含来自单名参与者的 $n = 12$ 个试次、 $m = 1049$ 个单元（V1区的体素）以及 $T = 100$ 张图像。
- **THINGS EEG2：** 一个脑电图数据集（Gifford 等，2022），其中10名参与者观看了来自 THINGS 数据库（Hebart 等，2019）的图像序列。预处理后的数据包含来自17个通道的信号，采样率为100 Hz。每名受试者进行了4次会话，每次会话包含同样200张图像的20次重复呈现。我们仅使用第一名参与者的数据，并将通道与时间点的每种组合作为一个独立单元。所得数据集包含来自单名参与者的 $n = 80$个试次、 $m = 1700$ 个单元（通道$\times$ 时间点）以及 $T = 200$ 张图像。
- **Sherlock数据集：**一个功能性磁共振成像数据集（Chen 等，2017 年），其中 16 名参与者观看了 BBC 电视连续剧《Sherlock》的第一集（48 分钟），同时接受 3T 磁共振成像扫描仪扫描。我们使用 fMRIPrep（Esteban 等，2019 年）对数据进行了预处理，并将其映射到 MNI 空间。我们使用 Harvard-

牛津图谱分离出 V1区和 IT区脑区。总之，该数据集包含 $n = 16$ 跨被试试次，每个试次包含 $m = 623$单元（体素，V1区 296 个，IT区 327 个），序列长度为 $T = 1976$ 时间点。

Sherlock数据集显然处于异质状态，因为其试次来自不同的受试者，而这些受试者的大脑在解剖结构和功能上存在差异。THINGS fMRI1 和 EEG2 数据集的情况则较为模糊。它们的试次来自同一参与者的重复记录，但分布在多个会话中，这可能会引入一定的异质性（例如，不同会话之间的电极位置或内部状态不同）。不过，在生成模拟这两个数据集的合成数据时，我们仍假设其具有同质性，并在第 4 节讨论可能存在异质性的后果。

合成脑记录 我们构建了由 n 个试次和排列在二维网格上的 $m = 10 \times 10$ 个体素组成的合成数据集，这些数据根据公式（3）的信号模型生成。试次生成过程如图 1 所示。共享信号 $\boldsymbol{\phi}_k$ 可以通过两种不同的方式构建。为了模拟对应于一个持续时间较长的连续刺激的功能性磁共振成像 BOLD 信号（Sherlock数据集），我们根据泊松过程（$\lambda = 0.1$）模拟神经事件，其持续时间服从指数分布（$\mu = 5$），从而生成箱形时间序列。随后，将该时间序列与双伽马血流动力学响应函数（HRF）进行卷积，以生成时间分辨率为 1.5 秒的类BOLD信号。对于向受试者呈现一组在每个试次中重新打乱顺序的离散刺激的实验设置（THINGS fMRI1 和 THINGS EEG2），信号 $\boldsymbol{\phi_k}$ 生成方式为独立同分布的高斯过程。噪声项 $\boldsymbol{\eta}_i^{(\ell)}$ 由两个成分组成。第一项遵循与信号项具有相同的结构：它是试次特异性信号的线性组合 $\tilde{\boldsymbol{\phi}}_{k'}^{(\ell)}$，使用与共享信号相同的过程为每个试次独立生成，并采用试次特异性的混合权重 $\tilde{w}_{ik'}^{(\ell)}$。所有信号（共享信号和试次特异性信号）中共享信号所占的比例由参数 $\rho_{\text{shared}} = K/(K + K')$控制。第二个噪声项是白高斯噪声，其标准差在不同体素间变化，为 $\sigma_i = \sigma_0\, e^{sZ_i}$，其中 $Z_i \sim \mathcal{N}(0, 1)$，因此$\sigma_0$ 是噪声水平中位数，而 s 控制其在不同体素间的分布范围。混合权重的空间结构受到 Erhardt 等人（2012）的启发：对于每个信号 $\boldsymbol{\phi}_k$，在网格上均匀抽取一个中心坐标（x_k，y_k），并将体素 i 的权重 w_{ik} 设为高斯核在体素中心与（x_k，y_k）之间距离处的取值，其标准差为 σ_w。在同质状态下，各试次的坐标保持不变，从而得到常数权重 w_{ik}。在异质状态下，每个试次都会将中心坐标重新抽样为（$x_k^{(\ell)}$，$y_k^{(\ell)}$）=（x_k，y_k）+（δ_x，δ_y），其中（δ_x，δ_y）$\sim \mathcal{U}$（$[-\Delta, \Delta]^2$）在每个试次中独立抽取，从而产生围绕共同均值变化的试次特异性权重 $w_{ik}^{(\ell)}$。通过改变参数 ρ_{shared}、σ_0、s、σ_w和 Δ，可以对不同的实验设置和脑区进行建模。

有限样本T 偏差 图 2 展示了三个数据集及旨在模拟它们的合成数据集中的估计量随试次次数 n 变化的情况。对于合成数据，可以将经验值与其闭式表达式的主导项进行比较；实线与对应黑色点线之间的差距可用于估计该估计量的偏差 $\mathbb{E}[\varepsilon_T]$。除了在较小 n下的 ATV 和 SHSB 外，两条线都十分接近；在这种情况下，当噪声上限较低时会出现正偏差。这就是第 3.2 节所述的瞬态区域，它出现在真实噪声上限较低的体素中。随着真实噪声上限随 n增加，偏差在较大的 n下消失。通过在固定 n下绘制估计量随 T变化的情况，可以进一步研究这一现象，附录 D 中对Sherlock数据集进行了这样的分析。

同质状态 观察图 2 的前两行可见，ATV 和 SHSB 的主导项都与 NC^{mean}重合，但存在有限样本-T偏差。在这里，由于合成数据中的所有试次都具有相同的噪声方差，因此 $\Delta_{AB}^{\text{sd}} = 0$，使得 ATV 和 SHSB 相等。在脑数据中，ATV 和 SHSB 也几乎重合，表明 Δ_{AB}^{sd} 在实践中很小。对于试次特定噪声上限，随着 n 增加，LOOC 和 TTMC 分别从下方和上方正确地趋近于 NC^{single}。

异质状态 在图 2 的底部两行中，ATV 和 SHSB 低估了 NC^{mean}，误差项的阶为 $1/n$，这与它们在异质状态下的闭式形式一致。该项会随试次间权重方差的增大而增大，而合成数据通过 Δ 和 σ_w控制这一方差。试次特定估计量 LOOC 和 TTMC 收敛到一个低于 NC^{single}的共同极限。因此，TTMC 不再是上界，因为在两种合成数据设置中，它都低于来自 $n \approx 8$ 的 NC^{single}。在两个脑区中，LOOP 都高于 LOOC。然而，在合成数据中，它趋于远低于

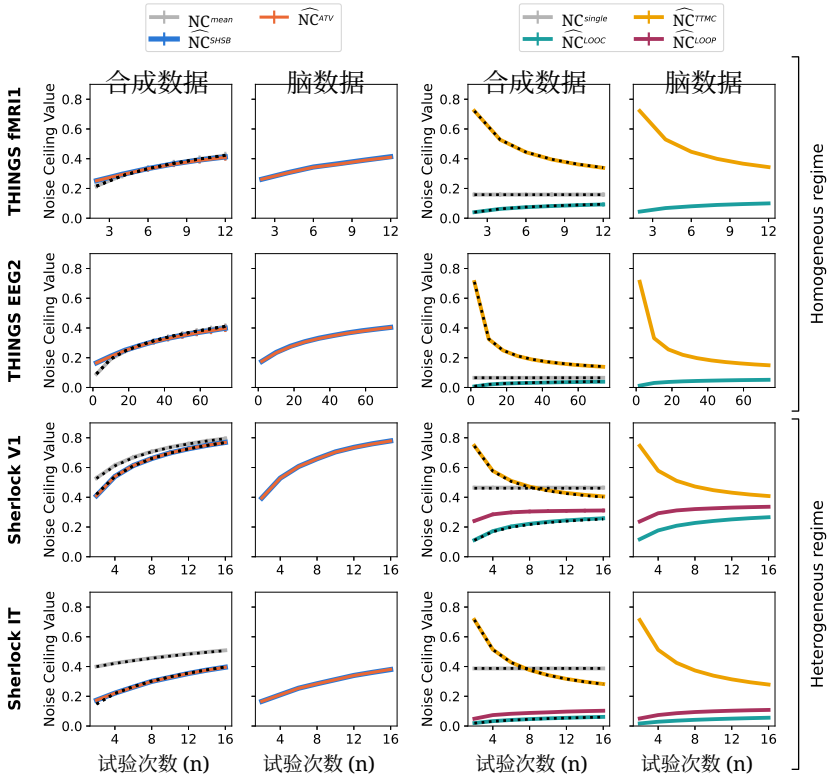

图 2：**合成数据和脑数据上的噪声上限估计量**。每一行对应第 3.4 节中的一个数据集。左列显示试次平均噪声上限 NC^{mean} 及其估计量，右列显示单试次噪声上限 NC^{single} 及其估计量，二者均为试验次数 n的函数。在每一列中，合成数据和脑数据并排显示，所有数值均在单元 $i = 1,\ldots,m$ 上取平均值。真实噪声上限（灰色）仅对合成数据已知。对于合成数据，实线表示 10 个模拟数据集的平均值，误差条表示渐近 95% 置信区间（通常小于线宽），黑色点线表示第 3.2、3.3 节中相应闭式形式的首项。合成数据在前两行中以同质状态（$\Delta = 0$）生成，在后两行中以异质状态（$\Delta > 0$）生成；参数经过选择，以模拟每行对应的脑数据（所有参数见表 4、附录 C）。对于每个 n，估计值均在 n 试次的 50 个随机子集上取平均。LOOP 没有闭式形式，因此仅针对 Sherlock 及其合成对应数据集显示。THINGS fMRI1 和 THINGS EEG2 的刺激过少（100 和 200 个），无法有意义地拟合岭回归；在实践中，LOOP 用于跨被试重复。

NC^{single}，因此，使用指数拟合将其外推到较大的 n 时，仍会低估噪声上限。合成数据的真实噪声上限不应被视为脑数据真实噪声上限的估计值。许多参数组合都可以在真实噪声上限不同的情况下重现相同的估计量曲线。原则上，NC^{mean} 可以位于 ATV 和 1 之间的任意位置，具体取决于异质性水平。合成数据的真实噪声上限只是脑数据真实噪声上限可能所在位置的一个合理示例。同样，LOOP 高于 LOOC 也不能证明存在异质性。在同质状态下，当相邻单元的信号存在相关性时，LOOP 可以结合许多单元，从而比 LOOC 更充分地平均掉它们的噪声。

4 讨论

在本研究中，我们为跨模态脑记录提出了统一的数学框架，并利用该框架分析了不同实验设置下噪声上限估计量的统计性质。我们确定了两种核心状态：同质状态，即所有试次都以相同的混合权重表达潜在信号；以及异质状态，即这些权重在不同试次之间有所变化。在同质状态下，ATV 可以恢复 NC^{mean}，误差项为有限-T。当所有试次具有相同噪声水平时，SHSB 对偶数个 n 也能做到这一点。随着 n 增大，LOOC 和 TTMC 收敛于 NC^{single}。在异质状态下，ATV 和 SHSB 会低估 NC^{mean}，低估量的阶为

表 2：推荐的噪声上限估计量。

	试次平均（NC^{mean}）	单试次（NC^{single}）
同质	ATV：在有限-T效应下无偏	LOOC：下界，TTMC：上界†
异质	ATV：下界	LOOC、LOOP 及 LOOP 变体中的最高值（在验证集上选择）：下界

†要求试次的噪声方差至少为平均噪声方差 $\sigma_\ell^2 \geq \bar{\sigma}^2/2$ 的一半。

$1/n$；LOOC 和 TTMC 则收敛于低于 NC^{single}的极限。在该状态下，LOOP 是唯一能够收敛于 NC^{single}的估计量，但对于现实中的 T值，其泛化误差会随预测变量数量增加而增大；在我们的合成数据中，它最终饱和于低于 NC^{single}的水平。当试次对应不同被试时，异质性通常十分重要，但它也可能出现在同一受试者的不同会话之间，甚至出现在同一会话内部：对首次看到的刺激所产生的反应，可能不同于对之后再次呈现、已经变得熟悉的刺激所产生的反应。这些因素会引入多大程度的异质性，是未来研究中的一个重要问题。如果异质性不可忽略，那么即使对于同一参与者中包含重复呈现的数据集，例如 THINGS-fMRI1 和 THINGS-EEG2，这些估计量也会低估真实噪声上限；不过，对于 ATV 和 SHSB，这种低估会随着重复次数增加而减小。

解析结果依赖于四个假设（假设 1–4）。试次之间存在相关性的噪声违反了第一个假设，并会使所有估计量偏大，因为这些估计量会将共享噪声视为信号。刺激锁定的非神经元反应，例如在特定事件发生时心率或注视的变化，并不违反该假设。它们在各试次之间具有可重复性，因此属于信号的一部分，原则上可以由刺激模型预测。对于在同一次扫描运行中记录的试次，共享噪声可能来自缓慢漂移，而高通滤波通常会将其去除。第二个假设，即信号具有有界性，对于任何物理记录而言都是显然成立的。最后两个假设要求噪声及其平方幅度具有短记忆，从而使误差项在 T 增大时消失。缓慢漂移和持续振荡会违反这两个假设，并可能导致误差缓慢衰减，甚至完全不衰减。通过高通滤波和干扰回归可以在很大程度上去除它们；剩余的记忆则可以通过观察信号及其平方的自相关衰减情况进行经验评估。

表 2 总结了应根据所处状态，以及模型是在试次平均响应还是单试次响应上进行评估，来选择使用哪种估计量。在同质状态下，可以在不存在系统误差的情况下估计试次平均噪声上限；而在异质状态下，只能得到下界。特别是，在异质状态下不应将 TTMC 用作上界。对于单试次响应，也可以使用 LOOP 的变体来降低泛化误差，但代价是更大的夹角 $\alpha^{(\ell)}$；例如，仅根据其他试次中的同一单元来预测每个单元，或根据其他试次的平均值进行预测。对于每个单元，取 LOOC、LOOP 及其变体中的最高值，可以为 NC^{single} 提供更紧的下界，前提是在验证集上选择估计量，以避免使估计值产生向上的偏差。当只能获得噪声上限的下界时，我们建议将脑分数报告为一个区间：区间下限为未归一化的脑分数，上限为以该下界归一化后的脑分数，并将其截断为不超过 1。附录 E 提供了计算噪声上限时常见问题的检查清单。

局限性 尽管我们的框架为噪声上限估计提供了统一的理论基础，但仍存在若干理论和实证方面的局限性。首先，可重复响应在形式上被定义为包含给定试次内固定属性后所有可预测的内容。这一选择与将编码模型作为记录响应预测变量的标准用法相一致。然而，如果目标是评估某个模型是否类似大脑，即它是否捕捉到了大脑之间的共同特征，那么应仅考虑大多数被试所共享的潜在信号，这将降低真实噪声上限。不过，本文的主要结论仍然成立，因为被试混合这些共享信号的方式仍会存在差异，从而导致估计量产生向下偏差。此外，尽管我们对留一预测（LOOP）提供了初步的理论分析，但完整的理论处理仍有待开展。在实证方面，我们的评估仅限于使用视觉（THINGS fMRI 和 THINGS EEG）或视听（Sherlock）刺激的功能性磁共振成像和脑电图数据集，因此尚未考察其他模态（侵入式电生理学、脑磁图）以及其他认知领域（语言处理、运动控制）。最后，我们的实证验证依赖于：使用在我们理论框架的某一特定情形下生成的合成数据，对闭式表达式进行一致性检验；以及在三个数据集之间进行定性比较，而不是在多样化的实验和合成设置下开展全面的定量基准测试。

致谢

本研究得到了瑞士国家科学基金会项目（编号：200021-236436）的支持。

参考文献

D. L. K. Yamins 和 J. J. DiCarlo，《使用目标驱动深度学习模型理解感觉皮层》，《自然-神经科学》，第19卷，第3期，第356–365页，2016年3月。 [在线]。可访问：http://dx.doi.org/10.1038/nn.4244 M. Schrimpf、J. Kubilius、M. J. Lee、N. A. Ratan Murty、R. Ajemian 和 J. J. DiCarlo，《通过综合基准测试推进人类智能的神经机制模型》，《神经元》，第108卷，第3期，第413–423页，2020年11月。 [在线]。可访问：http://dx.doi.org/10.1016/j.neuron.2020.07.040 A. Doerig、R. P. Sommers、K. Seeliger、B. Richards、J. Ismael、G. W. Lindsay、K. P. Kording、T. Konkle、M. A. J. van Gerven、N. Kriegeskorte 和 T. C. Kietzmann，《神经连接主义研究计划》，《自然综述：神经科学》，第24卷，第7期，第431–450页，2023年7月。[在线]。可访问：https://www.nature.com/articles/s41583-023-00705-w S. Sadeh 和 C. Clopath，《神经人工智能的兴起：连接神经科学与人工智能》，《自然综述：神经科学》，第26卷，第10期，第583–584页，2025年10月。 [在线]。可访问：https://www.nature.com/articles/s41583-025-00954-x M. Schrimpf、J. Kubilius、H. Hong、N. J. Majaj、R. Rajalingham、E. B. Issa、K. Kar、P. Bashivan、J. Prescott-Roy、F. Geiger、K. Schmidt、D. L. K. Yamins 和 J. J. DiCarlo，《Brain-Score：哪种人工神经网络最像大脑地进行物体识别？》，2018年9月，页码：407007，栏目：新结果。 [在线]。可访问：https://www.biorxiv.org/content/10.1101/407007v2 D. L. K. Yamins、H. Hong、C. F. Cadieu、E. A. Solomon、D. Seibert 和 J. J. DiCarlo，《性能优化的层级模型能够预测高级视觉皮层中的神经反应》，《美国国家科学院院刊》，第111卷，第23期，第8619–8624页，2014年6月。S.-M. Khaligh-Razavi 和 N. Kriegeskorte，《有监督但非无监督的模型或可解释IT皮层表征》，《公共科学图书馆·计算生物学》，第10卷，第11期，文章编号e1003915，2014年11月。 [在线]。可访问：https://journals.plos.org/ploscompbiol/article?id=10.1371/journal. pcbi.1003915 S. A. Cadena、G. H. Denfield、E. Y. Walker、L. A. Gatys、A. S. Tolias、M. Bethge 和 A. S. Ecker，《深度卷积模型改进了对自然图像中猕猴V1区反应的预测》，《公共科学图书馆·计算生物学》，第15卷，第4期，文章编号e1006897，2019年。 [在线]。可访问：https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006897 A. J. E. Kell、D. L. K. Yamins、E. N. Shook、S. V. Norman-Haignere 和 J. H. McDermott，《任务优化神经网络复现人类听觉行为、预测脑反应并揭示皮层加工层级》，《神经元》，第98卷，第3期，第630–644.e16页，2018年5月。[在线]。可访问：https://www.cell.com/neuron/abstract/S0896-6273(18)30250-2 G. Tuckute、J. Feather、D. Boebinger 和 J. H. McDermott，《许多但并非所有深度神经网络音频模型都能捕捉脑反应，并表现出模型阶段与脑区之间的对应关系》，《公共科学图书馆·生物学》，第21卷，第12期，文章编号e3002366，2023年12月。 [在线]。可访问：https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002366 L. Wehbe、A. Fyshe 和 T. M. Mitchell，《大脑中的语言加工：将神经活动映射到语言意义》，载于《人类语言：从基因和大脑到行为》，W. Zuidema 和 P. Hagoort 编。麻省理工学院出版社，2019年10月。 [在线]。可访问：https://doi.org/10.7551/mitpress/10841.003.0029 M. Schrimpf、I. A. Blank、G. Tuckute、C. Kauf、E. A. Hosseini、N. Kanwisher、J. B. Tenenbaum 和 E. Fedorenko，《语言的神经架构：综合建模趋向于预测性加工》，《美国国家科学院院刊》，第118卷，第45期，文章编号e2105646118，2021年11月。 [在线]。可访问：https://www.pnas.org/doi/10.1073/pnas.2105646118

C. Caucheteux、A. Gramfort 和 J.-R. King，《人类大脑听取语音时存在预测编码层级的证据》，*NatureHumanBehaviour*，第7卷，第3期，第430–441页，2023年3月。[在线]。可访问：https://www.nature.com/articles/s41562-022-01516-2

Y. Lei、X. Ge、Y. Zhang、Y. Yang 和 B. Ma，“大语言模型的思维方式是否类似于大脑？来自逐层嵌入和功能性磁共振成像的句子级证据”，*AAAI*人工智能会议论文集，第40卷，第1期，第579–587页，2026年3月。[在线]。可访问：https://ojs.aaai.org/index.php/AAAI/article/view/37022

B. AlKhamissi、G. Tuckute、Y. Tang、T. O. A. Binhuraib、A. Bosselut 和 M. Schrimpf，“从语言到认知：大语言模型如何超越人类语言网络”，载于*2025*年自然语言处理实证方法会议论文集，C. Christodoulopoulos、T. Chakraborty、C. Rose 和 V. Peng 编，苏州，中国：计算语言学协会，2025年11月，第24 321–24 339页。[在线]。可访问：https://aclanthology.org/2025.emnlp-main.1237/

R. Antonello、A. Vaidya 和 A. Huth，“功能性磁共振成像中的语言编码模型缩放定律”，载于《神经信息处理系统进展》，第 36 卷。Curran Associates, Inc.，2023 年，第 21 895–21 907 页。[在线]。可访问：https://proceedings.neurips.cc/paper_files/paper/2023/hash/4533e4a352440a32558c1c227602c323-Abstract-Conference.html

A. Gokce 和 M.Schrimpf，“面向任务优化的灵长类视觉腹侧通路模型的缩放定律”，2025年6月。[在线]。可访问：https://openreview.net/forum?id=WxY61MmHYo

A. Hsu、A. Borst 和 F. E. Theunissen，“量化神经反应的变异性及其在验证模型预测中的应用”，*Network: Computation in Neural Systems*，第 15 卷，第 2 期，第 91–109 页，2004 年 1 月。[在线]。可访问：https://doi.org/10.1088/0954-898X_15_2_002

M. Sahani 和 J. Linden，“听觉皮层反应有多线性？”载于 神经信息处理系统进展，第 15 卷。MIT Press，2002 年。[在线]。可访问：https://papers. nips.cc/paper_files/paper/2002/hash/7b4773c039d539af17c883eb9283dd14-Abstract.html

O. Schoppe、N. S. Harper、B. D. B. Willmore、A. J. King 和 J. W. H. Schnupp，“测量神经模型的性能”，计算神经科学前沿，第 10 卷，2016 年 2 月。[在线]。可访问：https://www.frontiersin.org/journals/computational-neuroscience/articles/10. 3389/fncom.2016.00010/full

A. Lage-Castellanos、G. Valente、E. Formisano 和 F. De Martino，“计算功能性磁共振成像响应的计算模型最大性能的方法”，《公共科学图书馆·计算生物学》，第 15 卷，第 3 期，文章 e1006397，2019 年 3 月。[在线]。可访问：https://pmc.ncbi.nlm.nih.gov/articles/PMC6426260/

S. van Bree、M. Styrnal 和 M. Hebart，“您的模型解释了多少方差？关于使用分半信度计算噪声上限的澄清说明”，2025 年 12 月。[在线]。可访问：https://osf.io/gjk45_v1 E. J. Allen、G. St-Yves、Y. Wu、J. L. Breedlove、J. S. Prince、L. T. Dowdle、M. Nau、B. Caron、F. Pestilli、I. Charest、J. B. Hutchinson、T. Naselaris 和 K. Kay，“用于连接认知神经科学与人工智能的大型 7T功能性磁共振成像数据集”，《自然-神经科学》，第 25 卷，第 1 期，第 116–126 页，2022 年 1 月。[在线]。可访问：https://www.nature.com/articles/s41593-021-00962-x M. Freteault、M. Le Clei、L. Tetrel、L. Bellec 和 N. Farrugia，“听觉人工网络与大规模个体功能性磁共振成像脑数据的对齐，可推广地改善脑编码和下游任务表现”，影像神经科学，第 3 卷，第 imag_a_00525 页，2025 年 4 月。[在线]。可访问：https://doi.org/10.1162/imag_a_00525J. Chen、Y. Leong、C. Honey、C. Yong、K. Norman 和 U. Hasson，“共享记忆揭示个体间神经活动的共享结构”，《自然-神经科学》，第 20 卷，第 1 期，第 115–125 页，2017 年 1 月。[在线]。可访问：https://pmc.ncbi.nlm.nih.gov/articles/PMC5191958/ Y. Tang、A. Gokce、K. J. Al-Karkari、D. Yamins 和 M. Schrimpf，“单一目标驱动下视觉通路中多样的知觉表征涌现”，2025 年 10 月，iSSN：2692-8205 页码：2025.07.22.664908 章节：新结果。[在线]。可访问：https://www.biorxiv.org/content/10.1101/2025.07.22.664908v3

G. Tuckute、A. Sathe、S. Srikant、M. Taliaferro、M. Wang、M. Schrimpf、K. Kay 和 E. Fedorenko，“使用大语言模型驱动和抑制人类语言网络”，《自然-人类行为》，第 8 卷，第 3 期，第 544–561 页，2024 年 3 月。 [在线]。可访问：https://www.nature.com/articles/s41562-023-01783-7

J. P. Cunningham 和 B. M. Yu，“大规模神经记录数据的降维”，《自然-神经科学》，第 17 卷，第 11 期，第 1500–1509 页，2014 年 11 月。 [在线]。可访问：https://www.nature.com/articles/nn.3776

P. Gao、E. Trautmann、B. Yu、G. Santhanam、S. Ryu、K. Shenoy 和 S. Ganguli，“多神经元维度、动力学与测量理论”，2017 年 11 月，页码：214262，章节：新结果。 [在线]。可访问：https://www.biorxiv.org/content/10.1101/214262v2

U. Hasson、Y. Nir、I. Levy、G. Fuhrmann 和 R. Malach，“自然视觉期间的被试间皮层活动同步”，《科学》，第 303 卷，第 5664 期，第 1634–1640 页，2004 年 3 月。 [在线]。可访问：https://www.science.org/doi/10.1126/science.1089506

E. Simony、C. J. Honey、J. Chen、O. Lositsky、Y. Yeshurun、A. Wiesel 和 U. Hasson，“叙事理解过程中默认模式网络的动态重构”，*NatureCommunications*，第 7 卷，第 1 期，第 12141 页，2016 年 7 月。 [在线]。可访问：https://www.nature.com/articles/ncomms12141 P.-H. C. Chen、J. Chen、Y. Yeshurun、U. Hasson、J. Haxby 和 P. J. Ramadge，“降维功能磁共振成像共享响应模型”，载于《神经信息处理系统进展》，第 28 卷。Curran Associates, Inc.，2015 年。 [在线]。可访问：https: //
p
roceedings.neurips.cc/paper/2015/hash/b3967a0e938dc2a6340e258630febd5a-Abstract.ht
ml J. V. Haxby、J. S. Guntupalli、S. A. Nastase 和 M. Feilong，“超对齐：对个体化皮层拓扑图中编码的共享信息进行建模”，《生命科学》，第 9 卷，第 e56601 页，2020 年 6 月。[在线]。可访问：https://doi.org/10.7554/eLife.56601 N. Kriegeskorte、M. Mur 和 P. Bandettini，“表征相似性分析——连接系统神经科学的各个分支”，*Frontiersin Systems Neuroscience*，第 2 卷，第 4 页，2008 年 11 月。[在线]。可访问：https://pmc.ncbi.nlm.nih.gov/articles/PMC2605405/ K. L. Aw、S. Montariol、B. AlKhamissi、M. Schrimpf 和 A. Bosselut，“指令微调使大语言模型与人脑对齐”，第一届语言建模会议，2024 年 8 月。 [在线]。可访问：https://openreview.net/forum?id=nXNN0x4wbl#discussion M. N. Hebart、O. Contier、L. Teichmann、A. H. Rockter、C. Y. Zheng、A. Kidder、A. Corriveau、M. Vaziri-Pashkam 和 C. I. Baker，“THINGS-data：用于研究人脑与行为中物体表征的大规模多模态数据集集合”，《生命科学》，2023 年 2 月。[在线]。可访问：https://elifesciences.org/articles/82580 M. N. Hebart、A. H. Dickter、A. Kidder、W. Y. Kwok、A. Corriveau、C. V. Wicklin 和 C. I. Baker，“THINGS：包含 1,854 个物体概念和超过 26,000 张自然情境物体图像的数据库”，《*PLOS ONE*》，第 14 卷，第 10 期，第 e0223792 页，2019 年 10 月。 [在线]。可访问：https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0223792 A. T. Gifford、K. Dwivedi、G. Roig 和 R. M. Cichy，“用于建模人类视觉物体识别的大型高丰富度脑电图数据集”，《神经影像》，第 264 卷，第 119754 页，2022 年 12 月。 [在线]。可访问：https://www.sciencedirect.com/science/article/pii/S1053811922008758 O. Esteban、C. J. Markiewicz、R. W. Blair、C. A. Moodie、A. I. Isik、A. Erramuzpe、J. D. Kent、M. Goncalves、E. DuPre、M. Snyder、H. Oya、S. S. Ghosh、J. Wright、J. Durnez、R. A. Poldrack 和 K. J. Gorgolewski，“fMRIPrep：用于功能磁共振成像的稳健预处理流程”，《自然·方法》，第 16 卷，第 1 期，第 111–116 页，2019 年 1 月。 [在线]。可访问：https://www.nature.com/articles/s41592-018-0235-4 E. B. Erhardt、E. A. Allen、Y. Wei、T. Eichele 和 V. D. Calhoun，“SimTB：基于时空可分离性模型的功能磁共振成像数据仿真工具箱”，《神经影像》，第 59 卷，第 4 期，第 4160–4167 页，2012 年 2 月。 [在线]。可访问：https://pmc.ncbi.nlm.nih.gov/articles/PMC3690331/

K. Garcia、E. McMahon、C. Conwell、M. Bonner 和 L. Isik，“动态社会视觉建模揭示深度学习与人类之间的差距”，*International Conference on Learning Representations*，第 2025 卷，第 71 337–71 362 页，2025 年 5 月。 [在线]。可访问：https://proceedings.iclr.cc/ paper_files/paper/2025/hash/b1bdb0f22c9748203c62f29aa297ac57-Abstract-Conference.html

A. LeBel、L. Wagner、S. Jain、A. Adhikari-Desai、B. Gupta、A. Morgenthal、J. Tang、L. Xu 和 A. G. Huth，“用于逐体素编码模型的自然语言功能磁共振成像数据集”，*Scientific Data*，第 10 卷，第 1 期，第 555 页，2023 年 8 月。 [在线]。可访问：https://www.nature.com/articles/s41597-023-02437-z

O. Moussa、D. Klakow 和 M. Toneva，“通过脑调优改善语音语言模型中的语义理解”，国际学习表征会议，第2025卷，第59 823–59 849页，2025年5月。 [在线]。可访问：https://proceedings.iclr.cc/paper_files/paper/2025/hash/95b7a93e60fdfd10cc202f44fd6adf5f-Abstract-Conference.html M. Khosla 和 L. Wehbe，“高级视觉区表现得像具有强选择性和功能专门化的领域通用滤波器”，*Nature Communications*，第17卷，第1期，第7484页，2026年6月。[在线]。可访问：https://www.nature.com/articles/s41467-026-73938-9 M. Honarmand、A. Sharma、B. AlKhamissi、J. Mehrer 和 M. Schrimpf，“在视觉语言模型中诱发阅读障碍”，国际学习表征会议，第2026卷，第45 253–45 273页，2026年4月。 [在线]。可访问：https://proceedings.iclr.cc/paper_files/paper/2026/hash/4b03821747e89ce803b2dac590f6a39b-Abstract-Conference.html M. Rahimi、Y. Yaghoobzadeh 和 M. R. Daliri，“大语言模型的解释揭示了大脑中的语言表征”，2025年4月，arXiv:2502.14671 [cs]。 [在线]。可访问：http://arxiv.org/abs/2502.14671 S. Kumar、T. R. Sumers、T. Yamakoshi、A. Goldstein、U. Hasson、K. A. Norman、T. L. Griffiths、R. D. Hawkins 和 S. A. Nastase，“基于Transformer的语言模型与人脑中的共享功能专门化”，*Nature Communications*，第15卷，第1期，第5523页，2024年6月。 [在线]。可访问：https://www.nature.com/articles/s41467-024-49173-5 G. Merlin 和 M. Toneva，“语言模型与大脑的对齐不仅源于下一词预测和词级信息”，载于*2024*年自然语言处理实证方法会议论文集，Y. Al-Onaizan、M. Bansal 和 Y.-N. Chen 编，美国佛罗里达州迈阿密：计算语言学协会，2024年11月，第18 431–18 454页。 [在线]。可访问：https://aclanthology.org/2024.emnlp-main.1024/ C. Kauf、G. Tuckute、R. Levy、J. Andreas 和 E. Fedorenko，“词汇-语义内容而非句法结构，是功能性磁共振成像响应在语言网络中的人工神经网络-脑相似性的主要贡献因素”，*Neurobiology ofLanguage*，第5卷，第1期，第7–42页，2024年4月。 [在线]。可访问：https://doi.org/10.1162/nol_a_00116E. A. Hosseini、M. Schrimpf、Y. Zhang、S. Bowman、N. Zaslavsky 和 E. Fedorenko，“人工神经网络语言模型即使在接受符合发展现实的训练量后，仍能预测人脑对语言的反应”，*Neurobiology of Language*，第5卷，第1期，第43–63页，2024年4月。 [在线]。可访问：https://direct.mit.edu/nol/article/5/1/43/119156/Artificial-Neural-Network-Language-Models-Predict Y.-C. Lin、H. Zhu 和 W. Schuler，“当控制维度扩展时，来自更大语言模型的向量对人类阅读时间和功能性磁共振成像数据的预测效果更差”，2025年5月，arXiv:2505.12196 [cs]。 [在线]。可访问：http://arxiv.org/abs/2505.12196 K. L. Aw 和 M. Toneva，“训练语言模型总结叙事文本可提高脑对齐”，第十一届国际学习表征会议，2022年9月。[在线]。可访问：https://openreview.net/forum?id=KzkLAE49H9b S. R. Oota、E. Çelik、F. Deniz 和 M. Toneva，“语音语言模型缺乏重要的与大脑相关的语义”，载于 计算语言学协会第*62*届年会（第*1*卷：长篇论文），L.-W. Ku、A. Martins 和 V. Srikumar 编，泰国曼谷：计算语言学协会，2024年8月，第8503–8528页。[在线]。可访问：https://aclanthology.org/2024.acl-long.462/

B. AlKhamissi、G. Tuckute、A. Bosselut 和 M. Schrimpf，“通过浅层未训练多头注意力网络实现类脑语言处理”，2024年6月，arXiv:2406.15109 [cs]。 [在线]。可访问：http://arxiv.org/abs/2406.15109 C. Gao、Z. Ma、J. Chen、P. Li、S. Huang 和 J. Li，“提高大语言模型与人脑语言处理的对齐度”，*Nature ComputationalScience*，第5卷，第11期，第1080–1090页，2025年11月。 [在线]。可访问：https://www.nature.com/articles/s43588-025-00863-0

文献中使用的噪声上限方法

表3：神经影像学文献中使用的噪声上限方法总结，按方法类型分组。对于每篇论文，我们报告刺激模态、是否使用一般线性模型（GLM）提取每个体素的贝塔值、数据集是否包含参与者在同一刺激上的多次重复，以及模型是否拟合用于预测参与者平均值（而非分别预测各参与者）。

噪声上限方法	论文	模态	GLM	重复次数	预测平均值-age
SHSB	Tang 等（2025）	视频	No	是和否	Yes
	Garcia等人（2025）	视频	Yes	Yes	No
	Tuckute等人（2023）	音频	是和否	Yes	No
	LeBel 等（2023）	语言	Yes	Yes	No
	Moussa 等（2025）	语言	No	Yes	No
ATV	Allen 等（2022）（NSD 数据集）	图像	Yes	Yes	No
	Khosla 和 Wehbe（2026）（NSD 数据集）	图像	Yes	Yes	No
	Hebart 等（2023）	图像	Yes	Yes	No
	Tuckute 等（2024）	语言	Yes	No	Yes
	Antonello 等（2023）	语言	No	Yes	No
LOOC	Honarmand 等（2026）	语言	No	No	No
TTMC	Rahimi 等（2025）	语言	No	No	No
	Kumar 等（2024）	音频	No	No	No
LOOP	Schrimpf 等（2021）	语言	No	Yes	No
	Merlin 和 Toneva（2024）	语言	No	Yes	No
	Kauf等（2024）	语言	Yes	Yes	No
	Hosseini等（2024）	语言	No	Yes	No
	Lin等（2025）	语言	No	Yes	No
	Aw 和 Toneva（2022）	语言	No	No	No
	Aw 等（2024）	语言	No	是与否	No
	Oota等（2024年）	语言	No	No	No
	AlKhamissi等（2024年）	语言	No	是和否	是和否
	Gao等（2025年）	语言	No	No	No

B 数学推导

B.1 关于使用人工智能的声明

引理 2、定理 1、推论 1、引理 3、命题 2 和命题 3 的证明是在人工智能的协助下完成的。研究过程中使用了大语言模型（Claude Opus 5 和 5.5），与作者以迭代方式共同开展证明框架的构思和证明草稿的撰写。作者核查了这些证明的每一步，检查了每个论证，并手工重新完成了所有计算。此外，作者还在必要时对论证进行了详细阐述、澄清或简化。所有其他结果均由作者独立推导。对于这些结果，大语言模型仅用于写作协助和检查推导。作者对所有陈述和证明的正确性承担全部责任。

B.2 Large T 收敛率

在本节中，我们固定一个记录单元，并省略索引 i 以提高可读性。对于每个试次ℓ，我们将相应的脑记录定义为随机过程 $\{v_t^{(\ell)}\}_{t>0}$。每个试次都带有属性 $\theta^{(\ell)}$（例如受试者身份、实验条件）。所有结果均以属性 $\theta = \left(\theta^{(1)}, \ldots, \theta^{(n)}\right)$为条件进行表述，即所有期望 $\mathbb{E}[\cdot]$、方差 Var (·) 和协方差 Cov (·) 均默认以 θ 为条件。

定义1（信号和噪声）。 试次 ℓ 的无噪信号和噪声定义为

$$v_t^{*(\ell)} := \mathbb{E}\left[v_t^{(\ell)}\right], \qquad \eta_t^{(\ell)} := v_t^{(\ell)} - v_t^{*(\ell)}, \tag{9}$$

因此， $\mathbb{E}[\eta_t^{(\ell)}] = 0$ 按构造即成立。给定 θ 时，信号 $v_t^{*(\ell)}$ 是确定性的，所有随机性都存在于噪声中。我们记

$$\gamma_\ell(t,s) := \mathrm{Cov}\big(\eta_t^{(\ell)}, \eta_s^{(\ell)}\big) \tag{10}$$

试次 ℓ 的噪声自协方差为

$$\sigma_\ell^2 := \frac{1}{T}\sum_{t=1}^{T} \mathrm{Var}\big(\eta_t^{(\ell)}\big) = \frac{1}{T}\sum_{t=1}^{T} \gamma_\ell(t,t) \tag{11}$$

试次 ℓ 的噪声方差在时间上的平均值为

定义2 （时间平均）。 对于有限序列 $\mathbf{x} = \{x_t\}_{t=1,\dots,T}$、$\mathbf{y} = \{y_t\}_{t=1,\dots,T}$， 我们记

$$\langle \mathbf{x} \rangle := \frac{1}{T}\sum_{t=1}^{T} x_t, \qquad \widehat{\mathrm{Cov}}_t\left(\mathbf{x},\mathbf{y}\right) := \langle \mathbf{xy} \rangle - \langle \mathbf{x} \rangle \langle \mathbf{y} \rangle, \tag{12}$$

$$\widehat{\mathrm{Var}}_t\left(\mathbf{x}\right) := \widehat{\mathrm{Cov}}_t\left(\mathbf{x},\mathbf{x}\right), \qquad \widehat{\mathrm{Corr}}_t\left(\mathbf{x},\mathbf{y}\right) := \frac{\widehat{\mathrm{Cov}}_t\left(\mathbf{x},\mathbf{y}\right)}{\sqrt{\widehat{\mathrm{Var}}_t\left(\mathbf{x}\right)\widehat{\mathrm{Var}}_t\left(\mathbf{y}\right)}}, \tag{13}$$

其中$\mathbf{xy}$ 是逐元素乘积。

定义3 （试次子集上的平均）。 对于非空试次子集 $A \subseteq [n]$， 我们记

$$\mathbf{v}^A := \frac{1}{|A|}\sum_{\ell\in A} \mathbf{v}^{(\ell)}, \qquad \mathbf{v}^{*A} := \frac{1}{|A|}\sum_{\ell\in A} \mathbf{v}^{*(\ell)}, \qquad \boldsymbol{\eta}^A := \frac{1}{|A|}\sum_{\ell\in A} \boldsymbol{\eta}^{(\ell)} \tag{14}$$

使得 $\mathbf{v}^A = \mathbf{v}^{*A} + \boldsymbol{\eta}^A$。此外，对于两个子集 A， $B \subseteq [n]$ 我们将它们的共享噪声定义为

$$s\left(A,B\right) := \frac{1}{|A||B|}\sum_{\ell\in A\cap B} \sigma_\ell^2. \tag{15}$$

现在列出推导以下结果所需的全部假设。

假设1 （独立试次）。 噪声过程 $\boldsymbol{\eta}^{(1)}$、.....、 $\boldsymbol{\eta}^{(n)}$ 彼此相互独立（在给定 θ 的条件下）。

假设2（有界信号）。 存在一个常数 $M < \infty$， 使得 $\sup^{\ell\in[n]}\sup_{t>0}\left|v_t^{*(\ell)}\right| < M$.

假设3 （短程时间相关性）。 存在一个常数 $C < \infty$，使得 $\sup^{\ell\in[n]}\sup_{t>0}\sum_{s=1}^{\infty}|\gamma_\ell\left(t,\ s\right)| < C$。

假设4 （噪声平方的短程相关性）。 对于每个试次 ℓ，对所有 t 都有 $\mathbb{E}\left[\left(\eta_t^{(\ell)}\right)^4\right] < \infty$ ，并且存在一个常数 $K_4 < \infty$ ，使得 $\sup^{\ell\in[n]}\sup_{t>0}\sum_{s=1}^{\infty}\left|\mathrm{Cov}\left(\left(\eta_t^{(\ell)}\right)^2,\ \left(\eta_s^{(\ell)}\right)^2\right)\right| < K_4$。

误差项。 在全文中， ε_T表示一个随机误差项，其偏差和均方误差满足 $|\mathbb{E}\left[\varepsilon_T\right]| \le c/T$和 $\mathbb{E}\left[\varepsilon_T^2\right] \le c/T$； $\varepsilon_T^{\max}$表示一个满足 $|\mathbb{E}\left[\varepsilon_T^{\max}\right]| \le (c/T)^{1/4}$和 $\mathbb{E}\left[(\varepsilon_T^{\max})^2\right] \le (c/T)^{1/2}$的随机误差项，其中常数 c与 T和 n均无关。每项结果都会明确指出 c可能依赖的量。

引理1 （噪声统计量）。 设 A， $B \subseteq [n]$ 为非空集。在假设 1–4 下，

(a) **（噪声的时间平均）** $\mathbb{E}\left[\langle \boldsymbol{\eta}^A \rangle\right] = 0$ 和方差 $\left(\langle \boldsymbol{\eta}^A \rangle\right) \le \frac{C}{|A|T}$ (b) **（信号-噪声协方差）** $\mathbb{E}\left[\widehat{\text{协方差}}_t\left(\mathbf{v}^{*A},\ \boldsymbol{\eta}^B\right)\right] = 0$ 和方差 $\left(\widehat{\text{协方差}}_t\left(\mathbf{v}^{*A},\ \boldsymbol{\eta}^B\right)\right) \le \frac{4M^2C}{|B|T}$

（c）**（噪声-噪声协方差）**写作 $\widehat{\text{协方差}}_t\left(\boldsymbol{\eta}^A,\ \boldsymbol{\eta}^B\right)=s\left(A,\ B\right)+Y_{AB}+W_{AB}$ 其中

$$Y_{AB}:=\langle\boldsymbol{\eta}^A\boldsymbol{\eta}^B\rangle-s(A,B),\qquad W_{AB}:=-\langle\boldsymbol{\eta}^A\rangle\langle\boldsymbol{\eta}^B\rangle. \tag{16}$$

于是 $\mathbb{E}\left[Y_{AB}\right]=0$、$\mathbb{E}\left[Y_{AB}^2\right]\le\frac{C^2+K_4}{T}$ 和 $\mathbb{E}\left[|W_{AB}|\right]\le\frac{C}{T}$。

证明。我们首先计算两个在整个证明过程中都会用到的量。根据假设1，不同试次的噪声互不相关，因此对于任意 A，

$$\operatorname{Cov}\left(\eta_t^A,\eta_s^A\right)=\frac{1}{|A|^2}\sum_{\ell,\ell'\in A}\operatorname{Cov}\left(\eta_t^{(\ell)},\eta_s^{(\ell')}\right)=\frac{1}{|A|^2}\sum_{\ell\in A}\gamma_\ell(t,s), \tag{17}$$

因此，根据假设3，对于每个 t，

$$\sum_{s=1}^{T}\left|\operatorname{Cov}(\eta_t^A,\eta_s^A)\right|\le\frac{1}{|A|^2}\sum_{\ell\in A}\sum_{s=1}^{\infty}|\gamma_\ell(t,s)|\le\frac{C}{|A|}. \tag{18}$$

（a）期望的线性性意味着 $\mathbb{E}\left[\langle\boldsymbol{\eta}^A\rangle\right]=0$。此外，

$$\operatorname{Var}\left(\langle\boldsymbol{\eta}^A\rangle\right)=\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\operatorname{Cov}(\eta_t^A,\eta_s^A)\le\frac{1}{T^2}\sum_{t=1}^{T}\frac{C}{|A|}=\frac{C}{|A|\,T}. \tag{19}$$

（b）令 $c_t:=v_t^{*A}-\langle\mathbf{v}^{*A}\rangle$。根据时间平均的线性性，

$$\widehat{\operatorname{Cov}}_t(\mathbf{v}^{*A},\boldsymbol{\eta}^B)=\langle\mathbf{v}^{*A}\boldsymbol{\eta}^B\rangle-\langle\mathbf{v}^{*A}\rangle\langle\boldsymbol{\eta}^B\rangle=\frac{1}{T}\sum_{t=1}^{T}\left(v_t^{*A}-\langle\mathbf{v}^{*A}\rangle\right)\eta_t^B=\frac{1}{T}\sum_{t=1}^{T}c_t\,\eta_t^B, \tag{20}$$

由于 c_t 是确定性的，其均值为零。根据假设2，$|v_t^{*A}|\le M$ 和 $|\langle\mathbf{v}^{*A}\rangle|\le M$，因此 $|c_t|\le 2M$。利用（18），

$$\operatorname{Var}\Big(\frac{1}{T}\sum_t c_t\eta_t^B\Big)=\frac{1}{T^2}\sum_{t,s}c_tc_s\operatorname{Cov}(\eta_t^B,\eta_s^B)\le\frac{4M^2}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\operatorname{Cov}(\eta_t^B,\eta_s^B)\right|\le\frac{4M^2C}{|B|\,T}. \tag{21}$$

（c）我们将平均值的乘积展开为单个试次的乘积，

$$\langle\boldsymbol{\eta}^A\boldsymbol{\eta}^B\rangle=\frac{1}{|A|\,|B|}\sum_{\ell\in A}\sum_{\ell'\in B}\langle\boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell')}\rangle=\frac{1}{|A|\,|B|}\sum_{\ell\in A}\sum_{\ell'\in B}\left(\frac{1}{T}\sum_{t=1}^{T}\eta_t^{(\ell)}\eta_t^{(\ell')}\right). \tag{22}$$

对于 $\ell\ne\ell'$，由独立性和均值为零可得 $\mathbb{E}\left[\langle\boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell')}\rangle\right]=0$。对于 $\ell=\ell'$，$\mathbb{E}\left[\langle\boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell)}\rangle\right]=\frac{1}{T}\sum_t\gamma_\ell(t,t)=\sigma_\ell^2$。因此 $\mathbb{E}\left[\langle\boldsymbol{\eta}^A\boldsymbol{\eta}^B\rangle\right]=\frac{1}{|A||B|}\sum_{\ell\in A\cap B}\sigma_\ell^2=s(A,B)$，从而得到 $\mathbb{E}[Y_{AB}]=0$。

仍对于 $\ell\ne\ell'$，两个试次的独立性给出 $\mathbb{E}[\eta_t^{(\ell)}\eta_s^{(\ell)}{}^{(\ell')}\eta^{(\ell')}{}_t{}_s\ ]=\gamma_\ell(t,s)\gamma_{\ell'}(t,s)$。注意，根据柯西-施瓦茨不等式，$|\gamma_\ell(t,s)|\le\sqrt{\gamma_\ell(t,t)\gamma_\ell(s,s)}\le C$，因为 $\gamma_\ell(t,t)\le\sum_s|\gamma_\ell(t,s)|\le C$。因此

$$\operatorname{Var}\left(\langle\boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell')}\rangle\right)=\frac{1}{T^2}\sum_{t,s}\gamma_\ell(t,s)\gamma_{\ell'}(t,s)\le\frac{1}{T^2}\sum_{t=1}^{T}C\sum_{s=1}^{\infty}|\gamma_\ell(t,s)|\le\frac{C^2}{T}. \tag{23}$$

对于 $\ell=\ell'$，假设4给出

$$\operatorname{Var}\left(\langle\boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell)}\rangle\right)=\frac{1}{T^2}\sum_{t,s}\operatorname{Cov}\left(\left(\eta_t^{(\ell)}\right)^2,\left(\eta_s^{(\ell)}\right)^2\right)\le\frac{K_4}{T}. \tag{24}$$

利用这些结果，并且由于 C^2， $K_4 \le C^2 + K_4$，我们可以计算

$$\begin{aligned}\operatorname{Var}(Y_{AB}) = \mathbb{E}\left[Y_{AB}^2\right] &= \mathbb{E}\left[\left(\frac{1}{|A||B|}\sum_{\ell\in A}\sum_{\ell'\in B}\left(\langle \boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell')}\rangle - \mathbb{E}\left[\langle \boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell')}\rangle\right]\right)\right)^2\right] \\ &\le \left(\frac{1}{|A||B|}\sum_{\ell\in A}\sum_{\ell'\in B}\sqrt{\mathbb{E}\left[\left(\langle \boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell')}\rangle - \mathbb{E}\left[\langle \boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell')}\rangle\right]\right)^2\right]}\right)^2 \\ &= \left(\frac{1}{|A||B|}\sum_{\ell\in A}\sum_{\ell'\in B}\sqrt{\operatorname{Var}\left(\langle \boldsymbol{\eta}^{(\ell)}\boldsymbol{\eta}^{(\ell')}\rangle\right)}\right)^2 \le \frac{C^2+K_4}{T}, \quad (25)\end{aligned}$$

其中，第一个不等式使用了闵可夫斯基不等式。

最后，利用柯西-施瓦茨不等式、$\langle \boldsymbol{\eta}^A\rangle$ 和 $\langle \boldsymbol{\eta}^B\rangle$ 均值为零这一事实，以及 (a)，我们可以写成

$$\mathbb{E}\left[|W_{AB}|\right] = \mathbb{E}\left[\left|\langle \boldsymbol{\eta}^A\rangle\langle \boldsymbol{\eta}^B\rangle\right|\right] \le \sqrt{\mathbb{E}\left[\langle \boldsymbol{\eta}^A\rangle^2\right]\mathbb{E}\left[\langle \boldsymbol{\eta}^B\rangle^2\right]} = \sqrt{\operatorname{Var}\left(\langle \boldsymbol{\eta}^A\rangle\right)\operatorname{Var}\left(\langle \boldsymbol{\eta}^B\rangle\right)} \le \frac{C}{T}. \quad (26)$$

□

定义 4（邻域）。对于 $\mathbf{z} \in \mathbb{R}^d$，我们将 $|\mathbf{z}|$ 记作欧几里得范数。对于集合 $\mathcal{K} \subset \mathbb{R}^d$和 $\delta > 0$，我们记作 $\mathcal{K}_\delta := \{\mathbf{x} \in \mathbb{R}^d : \ |\mathbf{x} - \mathbf{z}| \le \delta$，其中某个 $\mathbf{z} \in \mathcal{K}\}$ 表示与 $\mathcal{K}$的距离至多为 δ 的点集。

引理 2（δ方法）。设 $\{\hat{\mathbf{Z}}_T\}_{T>0}$为随机向量，$\{\mathbf{z}_T\}_{T>0}$为确定性的向量，且属于$\mathbb{R}^d$，满足

$$\hat{\mathbf{Z}}_T - \mathbf{z}_T = \mathbf{Y}_T + \mathbf{W}_T, \qquad \mathbb{E}\left[\mathbf{Y}_T\right] = 0, \qquad \mathbb{E}\left[|\mathbf{Y}_T|^2\right] \le \frac{c_1}{T}, \qquad \mathbb{E}\left[|\mathbf{W}_T|\right] \le \frac{c_2}{T}, \quad (27)$$

对于某些常数 c_1， c_2。设 $\mathcal{K}$为一个紧集，使得对于所有 T都有$\mathbf{z}_T \in \mathcal{K}$，令 $\delta > 0$，并令F：$\mathbb{R}^d \to \mathbb{R}$ 在包含 $\mathcal{K}_\delta$的某个开集上二阶连续可微。由于 $\mathcal{K}_\delta$ 是紧集，

$$L_1 := \sup_{\mathbf{x}\in\mathcal{K}_\delta}|\nabla F(\mathbf{x})| < \infty, \qquad L_2 := \sup_{\mathbf{x}\in\mathcal{K}_\delta}||\mathbf{H}_F(\mathbf{x})||_{\mathrm{op}} < \infty, \quad (28)$$

其中，$\mathbf{H}_F$ 是 F的 Hessian 矩阵。进一步假设存在一个常数 U，使得 ^$|F(\mathbf{Z}_T)| \le U$几乎处处成立，并且 $|F(\mathbf{z}_T)| \le U$。于是

(a) **(bias)** $\left|\mathbb{E}\left[F(\hat{\mathbf{Z}}_T)\right] - F(\mathbf{z}_T)\right| \le \frac{c}{T}$,

(b) **(mean squared error)** $\mathbb{E}\left[\left(F(\hat{\mathbf{Z}}_T) - F(\mathbf{z}_T)\right)^2\right] \le \frac{c}{T}$,

其中，c 仅取决于 c_1、c_2、δ、U、L_1 和 L_2。

^证明。为便于阅读，我们省略下标 T，并将 $\mathbf{D} := \mathbf{Z} - \mathbf{z} = \mathbf{Y} + \mathbf{W}$ 记作。

我们首先定义事件 G，在该事件上，^$\mathbf{Z}$ 接近$\mathbf{z}$：$= \{|\mathbf{Y}| \le \delta/2\} \cap \{|\mathbf{W}| \le \delta/2\}$。注意，只要 G 成立，就有 $\left|\hat{\mathbf{Z}} - \mathbf{z}\right| \le \delta$。根据马尔可夫不等式，

$$\begin{aligned}\mathbb{P}(G^c) &\le \mathbb{P}\left(|\mathbf{Y}| > \frac{\delta}{2}\right) + \mathbb{P}\left(|\mathbf{W}| > \frac{\delta}{2}\right) = \mathbb{P}\left(|\mathbf{Y}|^2 > \frac{\delta^2}{4}\right) + \mathbb{P}\left(|\mathbf{W}| > \frac{\delta}{2}\right) \\ &\le \frac{4c_1}{\delta^2 T} + \frac{2c_2}{\delta T} =: \frac{c_3}{T}. \quad (29)\end{aligned}$$

^利用 G 可知，$\mathbf{Z}$ 与$\mathbf{z}$ 之间的线段位于以$\mathbf{z} \in \mathcal{K}$为中心、半径为 δ 的球内，因此也位于 $\mathcal{K}_\delta$内。于是，根据中值定理和泰勒定理，在 G上有：

$$\left|F(\hat{\mathbf{Z}}) - F(\mathbf{z})\right| \le L_1|\mathbf{D}|, \qquad F(\hat{\mathbf{Z}}) - F(\mathbf{z}) = \nabla F(\mathbf{z})\cdot\mathbf{D} + R, \qquad |R| \le \frac{L_2}{2}|\mathbf{D}|^2. \quad (30)$$

(a) 利用（30），我们将偏差划分为四项:

$$\begin{aligned}
&\mathbb{E}\left[F(\hat{\mathbf{Z}})\right]-F(\mathbf{z})=\mathbb{E}\left[\left(F(\hat{\mathbf{Z}})-F(\mathbf{z})\right)\left(\mathbb{1}_G+\mathbb{1}_{G^c}\right)\right]\\
&=\mathbb{E}\left[\left(\nabla F(\mathbf{z})\cdot\mathbf{D}+R\right)\mathbb{1}_G\right]+\mathbb{E}\left[\left(F(\hat{\mathbf{Z}})-F(\mathbf{z})\right)\mathbb{1}_{G^c}\right]\\
&=\underbrace{\mathbb{E}\left[\left(\nabla F(\mathbf{z})\cdot\mathbf{Y}\right)\mathbb{1}_G\right]}_{I_1}+\underbrace{\mathbb{E}\left[\left(\nabla F(\mathbf{z})\cdot\mathbf{W}\right)\mathbb{1}_G\right]}_{I_2}+\underbrace{\mathbb{E}\left[R\,\mathbb{1}_G\right]}_{I_3}+\underbrace{\mathbb{E}\left[\left(F(\hat{\mathbf{Z}})-F(\mathbf{z})\right)\mathbb{1}_{G^c}\right]}_{I_4}.
\end{aligned}\tag{31}$$

由于 $\mathbb{E}[\mathbf{Y}]=0$，我们有

$$I_1=\mathbb{E}\left[\left(\nabla F(\mathbf{z})\cdot\mathbf{Y}\right)\right]-\mathbb{E}\left[\left(\nabla F(\mathbf{z})\cdot\mathbf{Y}\right)\mathbb{1}_{G^c}\right]=-\mathbb{E}\left[\left(\nabla F(\mathbf{z})\cdot\mathbf{Y}\right)\mathbb{1}_{G^c}\right],\tag{32}$$

因此根据柯西-施瓦茨不等式和（29），

$$|I_1|\le\mathbb{E}\left[|\nabla F(\mathbf{z})|\,|\mathbf{Y}|\,\mathbb{1}_{G^c}\right]\le L_1\mathbb{E}\left[|\mathbf{Y}|\,\mathbb{1}_{G^c}\right]\le L_1\sqrt{\mathbb{E}\left[|\mathbf{Y}|^2\right]}\sqrt{\mathbb{P}\left(G^c\right)}\le\frac{L_1\sqrt{c_1c_3}}{T}.\tag{33}$$

接下来，$|I_2|\le L_1\mathbb{E}[|\mathbf{W}|]\le L_1c_2/T$。在 G 上，$|\mathbf{D}|^2\le2|\mathbf{Y}|^2+2|\mathbf{W}|^2\le2|\mathbf{Y}|^2+\delta|\mathbf{W}|$，因此根据（30）

$$|I_3|\le\frac{L_2}{2}\mathbb{E}\left[2|\mathbf{Y}|^2+\delta|\mathbf{W}|\right]\le\frac{L_2\left(c_1+\delta c_2/2\right)}{T}.\tag{34}$$

最后，由于 $|F(\hat{\mathbf{Z}})|\le U$ 几乎处处成立，且 $|F(\mathbf{z})|\le U$，$|I_4|\le2U\,\mathbb{P}(G^c)\le2Uc_3/T$。将四个界相加，

$$\left|\mathbb{E}\left[F(\hat{\mathbf{Z}})\right]-F(\mathbf{z})\right|\le\frac{L_1\sqrt{c_1c_3}+L_1c_2+L_2\left(c_1+\delta c_2/2\right)+2Uc_3}{T}.\tag{35}$$

（b）我们将 G 和 G^c 上的均方误差进行划分。在 G 上，（30）和 $|\mathbf{D}|^2\le2|\mathbf{Y}|^2+\delta|\mathbf{W}|$ 给出 $\left(F(\hat{\mathbf{Z}})-F(\mathbf{z})\right)^2\le L_1^2\left(2|\mathbf{Y}|^2+\delta|\mathbf{W}|\right)$。在 G^c 上，$\left(F(\hat{\mathbf{Z}})-F(\mathbf{z})\right)^2\le4U^2$。因此，利用（29），

$$\begin{aligned}
\mathbb{E}\left[\left(F(\hat{\mathbf{Z}})-F(\mathbf{z})\right)^2\right]&\le L_1^2\left(2\mathbb{E}\left[|\mathbf{Y}|^2\right]+\delta\,\mathbb{E}\left[|\mathbf{W}|\right]\right)+4U^2\,\mathbb{P}\left(G^c\right)\\
&\le\frac{L_1^2\left(2c_1+\delta c_2\right)+4U^2c_3}{T}.
\end{aligned}\tag{36}$$

□

定义 5（总体对应量）。对于每个非空子集 $A\subseteq[n]$，令$\mathbf{x}^A$ 表示记录平均值$\mathbf{v}^A$ 或其无噪声对应量$\mathbf{v}^{*A}$。我们将$\mathbf{x}^A$ 的时间统计量的总体对应量定义为

$$\overline{\mathrm{Cov}}\left(\mathbf{x}^A,\mathbf{x}^B\right):=\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^{*A},\mathbf{v}^{*B}\right)+\begin{cases}s(A,B) & \text{if }\mathbf{x}^A=\mathbf{v}^A\text{ and }\mathbf{x}^B=\mathbf{v}^B,\\ 0 & \text{otherwise,}\end{cases}\tag{37}$$

$$\overline{\langle\mathbf{x}^A\rangle}:=\langle\mathbf{v}^{*A}\rangle,\tag{38}$$

以及$\overline{\mathrm{Var}}\left(\mathbf{x}^A\right):=\overline{\mathrm{Cov}}\left(\mathbf{x}^A,\mathbf{x}^A\right)$。这些量是确定性的。

定理 1（一般极限定理）。令 $\hat{\mathbf{Z}}_T\in\mathbb{R}^d$ 收集有限个统计量，其中每个统计量的形式均为 $\widehat{\mathrm{Cov}}_t\left(\mathbf{x}^A,\mathbf{x}^B\right)$ 或 $\langle\mathbf{x}^A\rangle$，或者为$\mathbf{x}$ m 中任意数量的统计量 S_1、.....、S_m 的平均值。令$\mathbf{z}_T$收集其总体对应量（37）–（38），其中平均值的总体对应量是总体对应量的平均值。令 K、δ、F 和U 满足引理 2 的条件。在假设 1–4 下，

$$F(\hat{\mathbf{Z}}_T)=F(\mathbf{z}_T)+\varepsilon_T,\tag{39}$$

误差项的常数 c仅取决于 M、C、K_4、d、δ、U、L_1和 L_2，而不取决于 n，也不取决于平均统计量的数量 m。

证明。根据引理 2，只需将 $\hat{\mathbf{Z}}_T - \mathbf{z}_T = \mathbf{Y}_T + \mathbf{W}_T$ 分解为 $\mathbb{E}[\mathbf{Y}_T] = 0$、$\mathbb{E}\left[|\mathbf{Y}_T|^2\right] \le c_1/T$ 和 $\mathbb{E}[|\mathbf{W}_T|] \le c_2/T$。由于 $\mathbb{E}\left[|\mathbf{Y}_T|^2\right] = \sum_{j=1}^{d} \mathbb{E}\left[Y_j^2\right]$ 和 $\mathbb{E}[|\mathbf{W}_T|] \le \sum_{j=1}^{d} \mathbb{E}[|W_j|]$，因此只需分别对每个条目 j 进行处理。利用 $\mathbf{v}^A = \mathbf{v}^{*A} + \boldsymbol{\eta}^A$ 以及 Cov_t 的双线性性质，共有五种情况。

(1) **时间平均值** $\langle \mathbf{x}^A \rangle$。如果 $\mathbf{x}^A = \mathbf{v}^{*A}$ 与总体对应量之间的差为 0。如果 $\mathbf{x}^A = \mathbf{v}^A$ 差为 $\langle \boldsymbol{\eta}^A \rangle$，我们取 $Y_j = \langle \boldsymbol{\eta}^A \rangle, W_j = 0$。根据引理1(a)，$\mathbb{E}[Y_j] = 0$ 以及 $\mathbb{E}\left[Y_j^2\right] \le C/T$。

(2) **无噪声协方差** $\widehat{\mathsf{Cov}}_t\left(\mathbf{v}^{*A}, \mathbf{v}^{*B}\right)$。差为 0。

(3) **混合协方差** $\widehat{\mathsf{Cov}}_t\left(\mathbf{v}^{*A}, \mathbf{v}^{B}\right)$（或对称地）。差为 $\widehat{\mathsf{Cov}}_t\left(\mathbf{v}^{*A}, \boldsymbol{\eta}^{B}\right)$，我们将其取为 Y_j，其中 $W_j = 0$。根据引理1(b)，$\mathbb{E}[Y_j] = 0$ 以及 $\mathbb{E}\left[Y_j^2\right] \le 4M^2C/T$。

(4) **记录协方差** $\widehat{\mathsf{Cov}}_t\left(\mathbf{v}^{A}, \mathbf{v}^{B}\right)$。差值为

$$\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^{*A}, \boldsymbol{\eta}^{B}\right) + \widehat{\mathrm{Cov}}_t\left(\boldsymbol{\eta}^{A}, \mathbf{v}^{*B}\right) + Y_{AB} + W_{AB}, \tag{40}$$

其中，Y_{AB} 和 W_{AB} 在引理1(c)中定义。我们将 Y_j 取为前三项之和以及 $W_j = W_{AB}$。根据引理1(b)-(c)，$\mathbb{E}[Y_j] = 0$，由闵可夫斯基不等式得

$$\begin{aligned}
\sqrt{\mathbb{E}\left[Y_j^2\right]} &\le \sqrt{\mathbb{E}\left[\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^{*A}, \boldsymbol{\eta}^{B}\right)^2\right]} + \sqrt{\mathbb{E}\left[\widehat{\mathrm{Cov}}_t\left(\boldsymbol{\eta}^{A}, \mathbf{v}^{*B}\right)^2\right]} + \sqrt{\mathbb{E}\left[Y_{AB}^2\right]} \\
&\le 2\sqrt{\frac{4M^2C}{T}} + \sqrt{\frac{C^2 + K_4}{T}} \\
&= \frac{4M\sqrt{C} + \sqrt{C^2 + K_4}}{\sqrt{T}},
\end{aligned} \tag{41}$$

而 $\mathbb{E}[|W_j|] \le C/T$。

(5) **平均值** $\frac{1}{m}\sum_{i=1}^{m} S_i$ 形式为(1)-(4)的统计量。S_i 与其总体对应量之间的每个差值都可分解为 $Y_{(i)} + W^{(i)}$，如上所述。我们取 $Y_j := \frac{1}{m}\sum_i Y_{(i)}$ 以及 $W_j := \frac{1}{m}\sum_i W^{(i)}$。于是 $\mathbb{E}[Y_j] = 0$，根据闵可夫斯基不等式和三角不等式，

$$\sqrt{\mathbb{E}\left[Y_j^2\right]} \le \frac{1}{m}\sum_{i=1}^{m}\sqrt{\mathbb{E}\left[\left(Y^{(i)}\right)^2\right]}, \qquad \mathbb{E}[|W_j|] \le \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}\left[\left|W^{(i)}\right|\right], \tag{42}$$

因此，情形（1）至（4）的界也对 Y_j 和 W_j 成立，且与 m 无关。

因此，引理 2 的条件对 $c_1 = d\left[\left(4M\sqrt{C} + \sqrt{C^2 + K_4}\right)^2 + C\right]$ 和 $c_2 = dC$ 成立，它们不依赖于 n，由引理 2 即可得到该结果。

□

推论 1（试次平均值的相关性）。设 A，$B \subseteq [n]$ 非空，并假设存在一个常数 $a_0 > 0$ 使得

$$\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*A}\right) \ge a_0, \qquad \widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*B}\right) \ge a_0 \qquad \text{for all } T. \tag{43}$$

在假设 – 下，1 4

$$\widehat{\mathrm{Corr}}_t\left(\mathbf{v}^{A}, \mathbf{v}^{B}\right) = \frac{\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^{*A}, \mathbf{v}^{*B}\right) + s(A,B)}{\sqrt{\left(\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*A}\right) + s(A,A)\right)\left(\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*B}\right) + s(B,B)\right)}} + \varepsilon_T, \tag{44}$$

$$\widehat{\mathrm{Corr}}_t\left(\mathbf{v}^{*A}, \mathbf{v}^{B}\right) = \frac{\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^{*A}, \mathbf{v}^{*B}\right)}{\sqrt{\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*A}\right)\left(\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*B}\right) + s(B,B)\right)}} + \varepsilon_T, \tag{45}$$

其中，在这两种情况下，误差项的常数 c 仅取决于 M、C、K_4 和 a_0。

证明。我们证明（44）；（45）的证明完全相同，只需将 $s(A,\ B)$ 和 $s(A,\ A)$ 设为零。我们应用定理 1，其中 $d=3$,

$$\hat{\mathbf{Z}}_T = \left(\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^A, \mathbf{v}^B\right), \widehat{\mathrm{Var}}_t\left(\mathbf{v}^A\right), \widehat{\mathrm{Var}}_t\left(\mathbf{v}^B\right)\right), \tag{46}$$

$$F(u_1,u_2,u_3) := \begin{cases} \frac{u_1}{\sqrt{u_2 u_3}} & \text{if } u_2>0 \text{ and } u_3>0, \\ 0 & \text{otherwise,} \end{cases} \tag{47}$$

从而 $F(\hat{\mathbf{Z}}_T) = \widehat{\text{相关性}}_t\left(\mathbf{v}^A,\ \mathbf{v}^B\right)$（约定当其中一个方差为零时，相关性为 0）。总体对应量为

$$\mathbf{z}_T = \Big(\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^{*A}, \mathbf{v}^{*B}\right) + s(A,B), \\ \widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*A}\right) + s(A,A),\ \widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*B}\right) + s(B,B)\Big), \tag{48}$$

且 $F(\mathbf{z}_T)$ 是（44）的首项。我们检查引理 2 的条件：

（a）**紧集** 根据假设2，$\widehat{\text{方差}}_t\left(\mathbf{v}^{*A}\right) = \langle\left(\mathbf{v}^{*A}\right)^2\rangle - \langle\mathbf{v}^{*A}\rangle^2 \le \langle\left(\mathbf{v}^{*A}\right)^2\rangle \le M^2$，对 B 也同样成立，因此根据柯西-施瓦茨不等式 $\left|\widehat{\text{协方差}}_t\left(\mathbf{v}^{*A},\ \mathbf{v}^{*B}\right)\right| \le M^2$。此外，$\sigma_\ell^2 \le C$（它是 $\gamma_\ell(t,\ t) \le C$ 的平均值），因此 $0 \le s(A,\ \underline{B}) \le \frac{|A\cap B|}{|A||B|} C \le C$。结合（43），$\mathbf{z}_T$ 位于紧集

$$\mathcal{K} := \left[-M^2 - C,\, M^2 + C\right] \times \left[a_0,\, M^2 + C\right] \times \left[a_0,\, M^2 + C\right]. \tag{49}$$

(b) **光滑性** 我们取 $\delta := a_0/2$。$\mathcal{K}_\delta$ 中的每个点（$u_1,\ u_2,\ u_3$）都满足$u_2 \ge a_0/2$ 和 $u_3 \ge a_0/2$，并且 F 在包含 $\mathcal{K}_\delta$的开集{（$u_1,\ u_2,\ u_3$）上无穷次可微：$u_2 > a_0/4,\ u_3 > a_0/4$}。

(c) **有界性** $|F(\hat{\mathbf{Z}}_T)| \le 1$，因为它是皮尔逊相关。对于总体对应量，令 τ在 $[T]$ 上均匀分布，并令 ξ_1、.....、ξ_n 为相互独立的中心化随机变量，且 $\mathrm{Var}(\xi_\ell) = \sigma_\ell^2$，并独立于 τ。令 $\xi^A := \frac{1}{|A|}\sum_{\ell\in A}\xi_\ell$（$\xi^B$同理），设 $X := v_\tau^{*A} + \xi^A$和 $Y := v_\tau^{*B} + \xi^B$。则 F（$\mathbf{z}_T$）$= \mathrm{Corr}(X,\ Y)$，因此 $|F$（$\mathbf{z}_T$）$| \le 1$。

引理 2 的所有条件均成立，定理 1 给出了该结果。□

引理 3（阈值函数）。设 $X_T = x_T + \varepsilon_T$，其中 x_T是确定性的，且 ε_T满足 $\mathbb{E}\left[\varepsilon_T^2\right] \le c/T$。则

$$\sqrt{\max\left(0, X_T\right)} = \sqrt{\max\left(0, x_T\right)} + \varepsilon_T^{\max}, \tag{50}$$

其中，$\varepsilon_T^{\max}$的常数是该假设中的常数 c。

证明。令 $f(x) := \sqrt{\max(0,x)}$，我们首先证明 $|f(x) - f(y)| \le \sqrt{|x-y|}$。如果 $x,\ y \le 0$，则差为 0；如果 $y \le 0 < x$，则 $|f(x) - f(y)| = f(x) = \sqrt{x} \le \sqrt{x-y}$（如果$x \le 0 < y$，则对称地同样成立）；如果 $x,\ y > 0$，则 $|f(x) - f(y)| = |\sqrt{x} - \sqrt{y}| \le \sqrt{(\sqrt{x}-\sqrt{y})^2} \le \sqrt{x+y-2\min(x,\ y)} = \sqrt{|x-y|}$。这给出

$$|\varepsilon_T^{\max}| = |f(X_T) - f(x_T)| \le \sqrt{|\varepsilon_T|}. \tag{51}$$

因此 $(\varepsilon_T^{\max})^2 \le |\varepsilon_T|$，从而 $\mathbb{E}[(\varepsilon_T^{\max})^2] \le \mathbb{E}[|\varepsilon_T|] \le \sqrt{\mathbb{E}[\varepsilon_T^2]} \le (c/T)^{1/2}$。此外，根据詹森不等式

$$|\mathbb{E}\left[\varepsilon_T^{\max}\right]| \le \mathbb{E}\left[|\varepsilon_T^{\max}|\right] \le \mathbb{E}\left[\sqrt{|\varepsilon_T|}\right] \le \sqrt{\mathbb{E}\left[|\varepsilon_T|\right]} \le \left(\mathbb{E}\left[\varepsilon_T^2\right]\right)^{1/4} \le (c/T)^{1/4}. \tag{52}$$

□

注释1（关于引理 3）。与引理 2 相比，引理 3 的收敛速度较慢，这是因为最大值函数在 0 处不可微。当序列 $\{x_T\}_{T>0}$ 有界且远离零时，便可回到引理 2 的情形，收敛速度也会更快。

B.3 无噪声信号的潜在表示

定义 6（潜在信号）。设 ϕ_1，……，ϕ_K 为长度为 T 的序列，这些序列关于时间均值为零且正交归一，

$$\langle \phi_k \rangle = 0, \qquad \langle \phi_j \phi_k \rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{otherwise}, \end{cases} \tag{53}$$

并假设每个试次的中心化无噪信号都可以写成

$$\mathbf{v}^{*(\ell)} - \langle \mathbf{v}^{*(\ell)} \rangle = \sum_{k=1}^{K} w_k^{(\ell)} \phi_k, \qquad w_k^{(\ell)} := \widehat{\mathrm{Cov}}_t \left(\mathbf{v}^{*(\ell)}, \phi_k \right). \tag{54}$$

对于非空子集 $A \subseteq [n]$，我们将 $\bar{w}_k^A := \frac{1}{|A|} \sum_{\ell \in A} w_k^{(\ell)}$ 记为平均权重，并将 $\bar{w}_k := \bar{w}_k^{[n]}$ 记为。

这种表示总是存在的：中心化无噪声信号 $\mathbf{v}^{*(\ell)} - \langle \mathbf{v}^{*(\ell)} \rangle$、$\ell \in [n]$ 张成了 $\mathbb{R}^T$ 的一个维数至多为 n 的子空间；对它们应用格拉姆-施密特正交化过程，并使用内积 $\langle \mathbf{xy} \rangle$，即可得到满足（54）的中心化正交归一序列。

引理4（无噪声协方差的潜在形式）。在定义 6 的表示下，对于任意非空 A，$B \subseteq [n]$，

$$\widehat{\mathrm{Cov}}_t \left(\mathbf{v}^{*A}, \mathbf{v}^{*B} \right) = \sum_{k=1}^{K} \bar{w}_k^A \bar{w}_k^B, \qquad \text{in particular} \qquad \widehat{\mathrm{Var}}_t \left(\mathbf{v}^{*A} \right) = \sum_{k=1}^{K} \left(\bar{w}_k^A \right)^2. \tag{55}$$

证明。对（54）关于 $\ell \in A$ 取平均可得 $\mathbf{v}^{*A} - \langle \mathbf{v}^{*A} \rangle = \sum_k \bar{w}_k^A \phi_k$，对 B 同理。由于 $\widehat{\text{协方差}}_t$ 是双线性的，并且在加上常数后保持不变，

$$\widehat{\mathrm{Cov}}_t \left(\mathbf{v}^{*A}, \mathbf{v}^{*B} \right) = \widehat{\mathrm{Cov}}_t \left(\sum_{j=1}^{K} \bar{w}_j^A \phi_j, \sum_{k=1}^{K} \bar{w}_k^B \phi_k \right) = \sum_{j=1}^{K} \sum_{k=1}^{K} \bar{w}_j^A \bar{w}_k^B \, \widehat{\mathrm{Cov}}_t \left(\phi_j, \phi_k \right) = \sum_{k=1}^{K} \bar{w}_k^A \bar{w}_k^B, \tag{56}$$

其中，最后一个等式使用了（53），该式给出 $\widehat{\mathrm{Cov}}_t \left(\phi_j, \phi_k \right) = \langle \phi_j \phi_k \rangle - \langle \phi_j \rangle \langle \phi_k \rangle = \langle \phi_j \phi_k \rangle$。

□

B.4 试次平均噪声上限

在本节中，我们推导试次平均响应 NC^{mean} 及其估计量 $\widehat{NC}^{\text{SHSB}}$ 和 $\widehat{NC}^{\text{ATV}}$ 的真实噪声上限的闭式表达式。

B.4.1 真实噪声上限 NC^{mean}

我们回顾一下，试次平均响应的真实噪声上限是无噪声平均值与记录平均值之间的相关性，$NC^{\text{mean}} := \widehat{\mathrm{Corr}}_t \left(\bar{\mathbf{v}}^*, \bar{\mathbf{v}} \right)$，其中 $\bar{\mathbf{v}}^* := \mathbf{v}^{*[n]}$，且 $\bar{\mathbf{v}} := \mathbf{v}^{[n]}$。我们记

$$\bar{\sigma}^2 := \frac{1}{n} \sum_{\ell=1}^{n} \sigma_\ell^2 \tag{57}$$

对于在试次和时间上取平均的噪声方差。

命题 1（真实噪声上限，试次平均响应）。假设存在一个常数 $a_0 > 0$，使得对于所有 T，$\mathrm{Var}_t \left(\bar{\mathbf{v}}^* \right) \geq a_0$。在假设 1–4 下，

$$NC^{\text{mean}} = \sqrt{\frac{\widehat{\mathrm{Var}}_t \left(\bar{\mathbf{v}}^* \right)}{\widehat{\mathrm{Var}}_t \left(\bar{\mathbf{v}}^* \right) + \bar{\sigma}^2 / n}} + \varepsilon_T = \sqrt{\frac{\sum_{k=1}^{K} \bar{w}_k^2}{\sum_{k=1}^{K} \bar{w}_k^2 + \bar{\sigma}^2 / n}} + \varepsilon_T, \tag{58}$$

其中，误差项的常数 c 仅取决于 M、C、K_4 和 a_0，而不取决于 n。第二个表达式在定义 6 的表示下成立。

证明。 我们将推论 1 中的（45）与 $A = B = [n]$ 一起应用。完整平均值与其自身的共享噪声为

$$s([n],[n]) = \frac{1}{n^2}\sum_{\ell=1}^{n}\sigma_\ell^2 = \frac{\bar\sigma^2}{n}, \tag{59}$$

$\widehat{\text{Cov}}_t(\bar{\mathbf{v}}^*,\ \bar{\mathbf{v}}^*) = \widehat{\text{Var}}_t(\bar{\mathbf{v}}^*)$。因此，推论 1 给出

$$NC^{\text{mean}} = \frac{\widehat{\text{Var}}_t(\bar{\mathbf{v}}^*)}{\sqrt{\widehat{\text{Var}}_t(\bar{\mathbf{v}}^*)\left(\widehat{\text{Var}}_t(\bar{\mathbf{v}}^*)+\bar\sigma^2/n\right)}} + \varepsilon_T = \sqrt{\frac{\widehat{\text{Var}}_t(\bar{\mathbf{v}}^*)}{\widehat{\text{Var}}_t(\bar{\mathbf{v}}^*)+\bar\sigma^2/n}} + \varepsilon_T, \tag{60}$$

这就是式（58）中的第一个表达式。第二个表达式由引理4 和 $A = [n]$，$\widehat{\text{Var}}_t(\bar{\mathbf{v}}^*) = \sum_k \bar w_k^2$ 得出。

□

B.4.2 跨试次方差方法 $\widehat{NC}^{\text{ATV}}$

我们假设在计算估计量之前，已对每个试次在时间维度上进行中心化，即将 $\mathbf{v}^{(\ell)}$ 替换为 $\mathbf{v}^{(\ell)} - \langle\mathbf{v}^{(\ell)}\rangle$。这是对每个扫描运行或受试者进行常规 z 标准化时的中心化步骤；不需要再除以标准差。使用每个时间步上试次间的无偏方差，

$$\widehat{\text{Var}}^{\text{u}}_\ell[v_t] := \frac{1}{n-1}\sum_{\ell=1}^{n}\left(v_t^{(\ell)} - \bar v_t\right)^2, \tag{61}$$

表 1 中的 ATV 估计量为

$$\widehat{NC}^{\text{ATV}} := \sqrt{\max\left(0,\hat F\right)}, \qquad \hat F := 1 - \frac{1}{n}\cdot\frac{\frac{1}{T}\sum_{t=1}^{T}\widehat{\text{Var}}^{\text{u}}_\ell[v_t]}{\widehat{\text{Var}}_t(\bar{\mathbf{v}})}, \tag{62}$$

约定当 $\widehat{NC}^{\text{ATV}} = 0$ if $\widehat{\text{Var}}_t(\bar{\mathbf{v}}) = 0$ 时。对于每个试次定义的量 $x^{(\ell)}$，我们将 $\widehat{\text{Var}}_\ell[x] := \frac{1}{n}\sum_{\ell=1}^{n}\left(x^{(\ell)} - \frac{1}{n}\sum_{\ell'=1}^{n}x^{(\ell')}\right)^2$ 记作其跨试次的经验方差。

命题 2 (ATV)。 假设存在一个常数 $a_0 > 0$，使得对于所有 T，$\widehat{\text{Var}}_t(\bar{\mathbf{v}}^*) \ge a_0$。在假设 1–4 下:

（a）**（一般情况）**

$$\widehat{NC}^{\text{ATV}} = \sqrt{\max\left(0, \frac{\sum_k \bar w_k^2 - \frac{1}{n-1}\sum_k \widehat{\text{Var}}_\ell[w_k]}{\sum_k \bar w_k^2 + \bar\sigma^2/n}\right)} + \varepsilon_T^{\max}; \tag{63}$$

（b）**（含信号单元）**如果存在 $f_{\min} > 0$，使得对于所有T，$\frac{\sum_k \bar w_k^2 - \frac{1}{n-1}\sum_k \widehat{\text{Var}}_\ell[w_k]}{\sum_k \bar w_k^2 + \bar\sigma^2/n} \ge f_{\min}$，则

$$\widehat{NC}^{\text{ATV}} = \sqrt{\frac{\sum_k \bar w_k^2 - \frac{1}{n-1}\sum_k \widehat{\text{Var}}_\ell[w_k]}{\sum_k \bar w_k^2 + \bar\sigma^2/n}} + \varepsilon_T. \tag{64}$$

误差项的常数 c 仅取决于 M、C、K_4、a_0（以及（b）中的 $f_{\min}$），而不取决于 n。

证明。 **(1) 重写估计量** 中心化后，对于所有 ℓ 和 $\langle\bar{\mathbf{v}}\rangle = 0$，$\langle\mathbf{v}^{(\ell)}\rangle = 0$，因此平方的时间平均值就是随时间变化的方差。由于 $\sum_\ell\left(v_t^{(\ell)} - \bar v_t\right)^2 = \sum_\ell\left(v_t^{(\ell)}\right)^2 - n\bar v_t^2$，

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\text{Var}}^{\text{u}}_\ell[v_t] = \frac{1}{n-1}\left(\sum_{\ell=1}^{n}\widehat{\text{Var}}_t\left(\mathbf{v}^{(\ell)}\right) - n\,\widehat{\text{Var}}_t(\bar{\mathbf{v}})\right). \tag{65}$$

因此 $\hat{F}=F\left(\hat{\mathbf{Z}}_T\right)$，其中

$$\hat{\mathbf{Z}}_T:=\left(\widehat{\mathrm{Var}}_t\left(\bar{\mathbf{v}}\right),\ \frac{1}{n}\sum_{\ell=1}^{n}\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{(\ell)}\right)\right),\qquad F(u_1,u_2):=1-\frac{u_2-u_1}{(n-1)u_1}. \tag{66}$$

$\hat{}$ 根据柯西-施瓦茨不等式，$\mathbf{Z}_T$ 的第二项至少与第一项一样大，因此得到 $F\le 1$。

(2) **总体对应量** $\hat{}\mathbf{Z}_T$ 的第一项是形如 $\widehat{\mathrm{Cov}}_t\left(\mathbf{x}^A,\mathbf{x}^B\right)$ 的统计量，第二项是 n 个统计量 $\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{(\ell)}\right)$ 的平均值。根据定义 5 和引理4，它们的总体对应量为

$$\begin{aligned}\mathbf{z}_T&:=\left(\sum_k\bar{w}_k^2+\frac{\bar{\sigma}^2}{n},\ \frac{1}{n}\sum_{\ell=1}^{n}\left(\sum_k\left(w_k^{(\ell)}\right)^2+\sigma_\ell^2\right)\right)\\&=\left(\sum_k\bar{w}_k^2+\frac{\bar{\sigma}^2}{n},\ \sum_k\bar{w}_k^2+\sum_k\widehat{\mathrm{Var}}_\ell\left[w_k\right]+\bar{\sigma}^2\right),\end{aligned} \tag{67}$$

其中我们使用了 $\frac{1}{n}\sum_\ell\left(w_k^{(\ell)}\right)^2=\bar{w}_k^2+\widehat{\mathrm{Var}}_\ell\left[w_k\right]$。因此

$$F\left(\mathbf{z}_T\right)=1-\frac{\frac{1}{n-1}\sum_k\widehat{\mathrm{Var}}_\ell\left[w_k\right]+\bar{\sigma}^2/n}{\sum_k\bar{w}_k^2+\bar{\sigma}^2/n}=\frac{\sum_k\bar{w}_k^2-\frac{1}{n-1}\sum_k\widehat{\mathrm{Var}}_\ell\left[w_k\right]}{\sum_k\bar{w}_k^2+\bar{\sigma}^2/n}. \tag{68}$$

(3) **紧集** 根据假设2，$\sum_k\bar{w}_k^2\le\frac{1}{n}\sum_\ell\sum_k w_k^{(\ell)2}\le\max_\ell\sum_k w_k^{(\ell)2}=\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*(\ell)}\right)\le M^2$。此外，根据假设3，$\bar{\sigma}^2\le C$。因此，$\mathbf{z}_T$ 属于 $\mathcal{K}_0$：$=\left[a_0,M^2+C\right]\times\left[a_0,M^2+C\right]$。另外，由于 $\frac{1}{n-1}\sum_k\widehat{\mathrm{Var}}_\ell\left[w_k\right]\le\frac{1}{n}\sum_\ell\sum_k\left(w_k^{(\ell)}\right)^2\le M^2$，$F(\mathbf{z}_T)\ge -M^2/a_0$。

(4) **光滑性** 我们写作 $F=1-\frac{1}{n-1}h$，其中 $h(u_1,\ u_2):=\frac{u_2-u_1}{u_1}$，且其不依赖于 n。h 在开集 $\{(u_1,\ u_2):\ u_1>a_0/4\}$上无穷次可微，该开集包含紧集 $(\mathcal{K}_0)^{a_0/2}$。因此，h 在 $(\mathcal{K}_0)^{a_0/2}$ 上一致连续（因为 $(\mathcal{K}_0)^{a_0/2}$ 是紧集），并且 F 在任意 $(\mathcal{K}_0)_\delta\subseteq(\mathcal{K}_0)^{a_0/2}$ 上无穷次可微，其一阶和二阶导数均独立于 $n\ge 2$ 有界，因为 $\frac{1}{n-1}\le 1$。

(5) (b) 的证明 令 $\mathcal{K}:=\{\mathbf{u}\in\mathcal{K}_0:\ F(\mathbf{u})\ge f_{\min}\}$，它是紧集并包含所有 $\mathbf{z}_T$，并选择 $\delta\le a_0/2$，使得当 $\mathbf{x}$、$\mathbf{y}\in(\mathcal{K}_0)^{a_0/2}$ 且 $|\mathbf{x}-\mathbf{y}|\le\delta$ 时，$|h(\mathbf{x})-h(\mathbf{y})|\le f_{\min}/2$（这是可能的，因为 h 一致连续）。由于 $\frac{1}{n-1}\le 1$、δ 不依赖于 n，且 $F\ge f_{\min}/2$ 依赖于 $\mathcal{K}_\delta$，因此 $G:=\sqrt{\max(0,F)}=\sqrt{F}$ 在包含 $\mathcal{K}_\delta$ 的开集上无穷次可微。此外，$0\le G\left(\hat{\mathbf{Z}}_T\right)\le 1$，因为 $\hat{}F\le 1$，并且 $0\le G(\mathbf{z}_T)\le 1$。将定理 1 应用于 G 即可得到 (b)。

~(6)（a）的证明 F 在 $u_1=0$ 附近不是有界的，因此我们先对其进行截断：令 $F:=\max\left(F,-1-M^2/a_0\right)$。令 $\mathcal{K}':=\left\{\mathbf{u}\in\mathcal{K}_0:\ F(\mathbf{u})\ge -M^2/a_0\right\}$，该集合是紧的，并且根据（3）包含所有 $\mathbf{z}_T$，再按照（5）选择 $\delta\le a_0/2$，但用 $1/2$ 替代 $f_{\min}/2$。在 $\mathcal{K}'_\delta$ 上，$F\ge -M^2/a_0-1/2$，因此 $\tilde{F}=F$ 在包含 $\mathcal{K}'_\delta$ 的某个开集上成立，并且在该开集上无穷次可微。此外，$\left|\tilde{F}\left(\hat{\mathbf{Z}}_T\right)\right|\le 1+M^2/a_0$ 且 $\left|\tilde{F}(\mathbf{z}_T)\right|\le 1+M^2/a_0$。将定理 1 应用于 $\tilde{F}$，得到 $\mathbb{E}\left[\left(\tilde{F}\left(\hat{\mathbf{Z}}_T\right)-F(\mathbf{z}_T)\right)^2\right]\le c/T$。由于 $\tilde{F}$ 和 F 在它们为正的地方重合，$\widehat{NC}^{\mathrm{ATV}}=\sqrt{\max\left(0,\tilde{F}\left(\hat{\mathbf{Z}}_T\right)\right)}$，再结合引理 3 以及 $X_T=\tilde{F}\left(\hat{\mathbf{Z}}_T\right)$ 和 $x_T=F(\mathbf{z}_T)$，即可得到（a）。

□

注 2. 试验间方差的归一化 如果使用归一化 $1/n$ 而不是 $1/(n-1)$ 来计算试验间方差，则相同的计算会在平方根下得到 $\sum_k \bar{w}_k^2 - \frac{1}{n}\sum_k \widehat{\mathrm{Var}}_{\ell}[w_k]+\bar{\sigma}^2/n^2$。因此会产生一个额外项 $\frac{1}{n-1}\frac{\sum_k \widehat{\mathrm{Var}}_{\ell}[w_k]+\bar{\sigma}^2/n^2}{\sum_k \bar{w}_k^2+\bar{\sigma}^2/n}$。

B.4.3 采用斯皮尔曼-布朗校正的分半相关 $\widehat{NC}^{\mathrm{SHSB}}$

在本节中，我们假设 $n \geq 2$ 为偶数。我们将一个<分法 id='2'>划分 定义为一对 (A, B) 不相交的子集，其中$A \cup B = [n]$ 且 $|A| = |B| = n/2$。对于一个划分，我们记作

$$\bar{\sigma}_A^2 := \frac{2}{n}\sum_{\ell \in A}\sigma_\ell^2 \tag{69}$$

对于 A 的噪声水平（B 同理），因此 $\frac{1}{2}\left(\bar{\sigma}_A^2 + \bar{\sigma}_B^2\right) = \bar{\sigma}^2$，且 $\hat{\rho}_{A,B} := \widehat{\mathrm{Corr}}_t\left(\mathbf{v}^A, \mathbf{v}^B\right)$ 表示两个半平均之间的相关性。表 1 中单个划分的 SHSB估计量为

$$\widehat{NC}_{A,B}^{\mathrm{SHSB}} := \sqrt{\max\left(0, \frac{2\hat{\rho}_{A,B}}{1+\hat{\rho}_{A,B}}\right)}. \tag{70}$$

在实践中，通常会对由 $\mathcal{S}$ 个 r 划分组成的集合（所有可能的划分，或 r随机选取的划分）计算该估计量并取平均，

$$\widehat{NC}^{\mathrm{SHSB}} := \sqrt{\max\left(0, \frac{1}{r}\sum_{(A,B)\in\mathcal{S}}\frac{2\hat{\rho}_{A,B}}{1+\hat{\rho}_{A,B}}\right)}. \tag{71}$$

命题 3 (SHSB)。 令 $(A,\ B)$ 为一个划分，并假设存在一个常数 $a_0 > 0$ 使得

$$\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*A}\right) \geq a_0, \qquad \widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*B}\right) \geq a_0, \qquad \widehat{\mathrm{Cov}}_t\left(\mathbf{v}^{*A}, \mathbf{v}^{*B}\right) \geq 0 \qquad \text{for all } T. \tag{72}$$

在假设 1–4 下:

(a) **（一般情况）**

$$\widehat{NC}_{A,B}^{\mathrm{SHSB}} = \sqrt{\frac{\sum_k \bar{w}_k^2 - \Delta_{AB}^w}{\sum_k \bar{w}_k^2 + \bar{\sigma}^2/n - \Delta_{AB}^{\mathrm{sd}}}} + \varepsilon_T^{\max}, \tag{73}$$

其中我们使用了

$$\Delta_{AB}^w := \frac{1}{4}\sum_k \left(\bar{w}_k^A - \bar{w}_k^B\right)^2,$$
$$\Delta_{AB}^{\mathrm{sd}} := \frac{1}{4}\left(\sqrt{\sum_k \left(\bar{w}_k^A\right)^2 + \frac{2\bar{\sigma}_A^2}{n}} - \sqrt{\sum_k \left(\bar{w}_k^B\right)^2 + \frac{2\bar{\sigma}_B^2}{n}}\right)^2; \tag{74}$$

(b) **（含信号单元）**如果存在 $f_{\min} > 0$，使得对于所有 T，(73) 中平方根内的参数至少为 $f_{\min}$，则 (73) 中的 $\varepsilon_T^{\max}$ 可替换为 ε_T；

(c) **（分半上的平均）**如果对于 $\mathcal{S}$ 的每个划分，(72) 都成立，

$$\widehat{NC}^{\mathrm{SHSB}} = \sqrt{\frac{1}{r}\sum_{(A,B)\in\mathcal{S}}\frac{\sum_k \bar{w}_k^2 - \Delta_{AB}^w}{\sum_k \bar{w}_k^2 + \bar{\sigma}^2/n - \Delta_{AB}^{\mathrm{sd}}}} + \varepsilon_T^{\max}. \tag{75}$$

此外，如果（75）中平方根的被开方数对于所有 T 至少为 $f_{\min} > 0$，则误差项变为 ε_T，如（b）所示。

常数 c 仅取决于 M、C、K_4、a_0（以及 (b)-(c) 中的 $f_{\min}$），而不取决于 n，也不取决于划分的数量 r。

证明。由于 $\hat{\rho}_{A,B} = u_1/\sqrt{u_2 u_3}$ 对于 $\hat{\mathbf{Z}}_T := \left(\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^A, \mathbf{v}^B\right), \widehat{\mathrm{Var}}_t\left(\mathbf{v}^A\right), \widehat{\mathrm{Var}}_t\left(\mathbf{v}^B\right)\right) =: (u_1, u_2, u_3)$，我们有 $\frac{2\hat{\rho}_{A,B}}{1+\hat{\rho}_A \hat{\rho}_{A,B}} = F\left(\hat{\mathbf{Z}}_T\right)$，其中

$$F(u_1, u_2, u_3) := \frac{2u_1}{u_1 + \sqrt{u_2 u_3}}. \tag{76}$$

这两组 A 和 B 不相交，因此 $s(A, B) = 0$，并且

$$s(A,A) = \frac{4}{n^2}\sum_{\ell \in A}\sigma_\ell^2 = \frac{2\bar{\sigma}_A^2}{n}, \qquad s(B,B) = \frac{2\bar{\sigma}_B^2}{n}. \tag{77}$$

因此，$\mathbf{Z}_T$ 的总体对应量为

$$\mathbf{z}_T = (z_1, z_2, z_3) := \left(\sum_k \bar{w}_k^A \bar{w}_k^B, \sum_k \left(\bar{w}_k^A\right)^2 + \frac{2\bar{\sigma}_A^2}{n}, \sum_k \left(\bar{w}_k^B\right)^2 + \frac{2\bar{\sigma}_B^2}{n}\right). \tag{78}$$

由于 $|A| = |B|$，$\bar{w}_k = \frac{1}{2}\left(\bar{w}_k^A + \bar{w}_k^B\right)$ 和 $\frac{1}{2}\left(\bar{\sigma}_A^2 + \bar{\sigma}_B^2\right) = \bar{\sigma}^2$，因此

$$z_2 + z_3 + 2z_1 = \sum_k \left(\bar{w}_k^A + \bar{w}_k^B\right)^2 + \frac{2\left(\bar{\sigma}_A^2 + \bar{\sigma}_B^2\right)}{n} = 4\left(\sum_k \bar{w}_k^2 + \frac{\bar{\sigma}^2}{n}\right). \tag{79}$$

利用这一点，我们将 $F(\mathbf{z}_T)$ 的分母改写为

$$\begin{aligned} z_1 + \sqrt{z_2 z_3} &= z_1 + \frac{1}{2}\left(z_2 + z_3 - \left(\sqrt{z_2} - \sqrt{z_3}\right)^2\right) \\ &= \frac{1}{2}\left(2z_1 + z_2 + z_3\right) - \frac{1}{2}\left(\sqrt{z_2} - \sqrt{z_3}\right)^2 \\ &= 2\left(\sum_k \bar{w}_k^2 + \frac{\bar{\sigma}^2}{n} - \Delta_{AB}^{\mathrm{sd}}\right). \end{aligned} \tag{80}$$

分子可以展开为

$$2\sum_k \bar{w}_k^A \bar{w}_k^B = \frac{1}{2}\sum_k \left(\left(\bar{w}_k^A + \bar{w}_k^B\right)^2 - \left(\bar{w}_k^A - \bar{w}_k^B\right)^2\right) = 2\left(\sum_k \bar{w}_k^2 - \Delta_{AB}^{w}\right). \tag{81}$$

结合 (81) 和 (80) 可知，$F(\mathbf{z}_T)$ 等于 (73) 中平方根内的表达式。为证明该命题，我们需要验证 F 的截断版本满足引理 2 的条件，然后使用 B.2 节的结果。

(1) 紧集、光滑性和有界性 如推论 1 的证明所示，$z_1 \le M^2$ 和 $z_2,\ z_3 \in [a_0, M^2 + C]$。此外，根据（72）有 $z_1 \ge 0$，并且根据柯西-施瓦茨不等式有 $z_1 = \sum_k \bar{w}_k^A \bar{w}_k^B \le \sqrt{\sum_k \left(\bar{w}_k^A\right)^2 \sum_k \left(\bar{w}_k^B\right)^2} \le \sqrt{z_2 z_3}$。因此，$\mathbf{z}_T$ 属于紧集 $\mathcal{K}$：$= [0, M^2] \times [a_0, M^2 + C] \times [a_0, M^2 + C]$，且 $0 \le F(\mathbf{z}_T) \le 1$。我们取 $\delta := a_0/8$，使得 $\mathcal{K}_\delta$ 中的每个点 $(u_1,\ u_2,\ u_3)$ 都满足 $u_1 \ge -a_0/8$ 和 $u_2,\ u_3 \ge 7a_0/8$，因而属于开集

$$\mathcal{O} := \left\{(u_1, u_2, u_3) : u_1 > -a_0/4,\ u_2 > 3a_0/4,\ u_3 > 3a_0/4\right\}. \tag{82}$$

在 $\mathcal{O}$，$u_1 + \sqrt{u_2 u_3} > a_0/2 > 0$，因此 F 在 $\mathcal{O}$ 上无穷次可微。此外，$F > -1$ 在 $\mathcal{O}$ 上：两边乘以正分母，$F(u_1, u_2, u_3) > -1$ 等价于 $3u_1 + \sqrt{u_2 u_3} > 0$，这成立是因为 $3u_1 > -3a_0/4$ 和 $\sqrt{u_2 u_3} > 3a_0/4$。为界定 $F\left(\hat{\mathbf{Z}}_T\right)$，我们定义 $\tilde{F} := \max(F, -1)$，其中 $\tilde{F} := 0$（若 $u_2 u_3 \le 0$），而 $\tilde{F} := -1$（若 $u_2 u_3 > 0$ 且 $u_1 + \sqrt{u_2 u_3} = 0$）。由此可得 $F = \tilde{F}$ 在 $\mathcal{O}$ 上成立，因此 $\tilde{F}$ 在包含 $\mathcal{K}_\delta$ 和 $\left|\tilde{F}(\mathbf{z}_T)\right| \le 1$ 的开集 $\mathcal{O}$ 上无穷次可微。此外，$\left|\tilde{F}\left(\hat{\mathbf{Z}}_T\right)\right| \le 1$，因为 $\hat{\rho}_{A,B} \in [-1, 1]$，且 $\frac{2\rho}{1+\rho} \le 1$ 对于 $-1 < \rho \le 1$ 成立。因此，$\tilde{F}$, $\mathcal{K}$, δ 和 $U = 1$ 满足引理 2 的条件。最后，由于 F 和 $\tilde{F}$ 在为正的区域上重合，

$$\widehat{NC}_{A,B}^{\mathrm{SHSB}} = \sqrt{\max\left(0, \tilde{F}\left(\hat{\mathbf{Z}}_T\right)\right)}.$$

˜(2)**定理 1 的 (a) 部分证明** 将定理 1 应用于 F，其中 $\mathcal{K}$, δ 和 $U = 1$，可得 $\mathbb{E}\left[\left(\tilde{F}\left(\hat{\mathbf{Z}}_T\right) - F(\mathbf{z}_T)\right)^2\right] \le c/T$；结合引理 3，其中 $X_T = \tilde{F}\left(\hat{\mathbf{Z}}_T\right)$ 和 $x_T = F(\mathbf{z}_T)$，即可得 (a)。

˜(3) (b) 的证明 令 $\mathcal{K} := \{\mathbf{u} \in \mathcal{K} : F(\mathbf{u}) \ge f_{\min}\}$，该集合是紧集，并包含所有 $\mathbf{z}_T$。由于 $\mathcal{K}_\delta$ 是 $\mathcal{O}$的紧子集，且 F 在其上连续可微，因此 $L := \tilde{\sup}_{\mathbf{u}\in\mathcal{K}_\delta} |\nabla F(\mathbf{u})| < \infty$。取 $\delta := \min(\delta, f_{\min}/(2L))$。对于 $\mathbf{y} \in \mathcal{K}_{\tilde{\delta}}$，存在˜ ˜$\mathbf{x} \in \mathcal{K} \subset \mathcal{K}$，使得 $|\mathbf{x} - \mathbf{y}| \le \delta \le \delta$。$\mathbf{x}$ 与 $\mathbf{y}$ 之间的线段位于 $\mathcal{K}_\delta$ 中，由中值定理可得˜值 $|F(\mathbf{y}) - F(\mathbf{x})| \le L\delta \le f_{\min}/2$，因此 $F(\mathbf{y}) \ge f_{\min}/2$。于是，在包含 $\tilde{\mathcal{K}_\delta}$的开集 $\mathcal{O} \cap \{F > f_{\min}/4\}$ 上，$G := \sqrt{\max\left(0, \tilde{F}\right)} = \sqrt{F}$是无穷次可微的。此外，根据步骤 (1)，$0 \le G\left(\hat{\mathbf{Z}}_T\right) \le 1$和 $0 \le G(\mathbf{z}_T) \le 1$。将定理 1˜ ˜应用于 G，取 $\mathcal{K}$、δ和 $U = 1$，即可得到 (b)。

（4）**（c）的证明** 令 $\mathcal{S} = \{(A_1, B_1) \ldots, (A_r, B_r)\}$，并令 $\hat{\mathbf{Z}}_T^{(j)}$ 和 $\mathbf{z}_T^{(j)}$ 为证明开头针对划分（A_j, B_j）所定义的向量，从而

$$\widehat{NC}^{\mathrm{SHSB}} = g\left(\frac{1}{r}\sum_{j=1}^{r} F\left(\hat{\mathbf{Z}}_T^{(j)}\right)\right), \qquad g(x) := \sqrt{\max(0, x)}, \tag{83}$$

˜我们分三步进行：首先，在每个划分中将 F 替换为其截断版本 $\tilde{F}$，然后对各个划分的展开式取平均值，最后应用 g。

（i）截断 令 $X := \frac{1}{r}\sum_j F\left(\hat{\mathbf{Z}}_T^{(j)}\right)$ 和 $\tilde{X} := \frac{1}{r}\sum_j \tilde{F}\left(\hat{\mathbf{Z}}_T^{(j)}\right)$。由于 $F \le \tilde{F}$ 和 g 在（$-\infty$, 1]）上是非递减的，并且取值于 $[0, \bar{1}]$，差值 $D_T := g(X) - g(\tilde{X})$ 满足 $-1 \le D_T \le 0$；除非对于某个j有 $F\left(\hat{\mathbf{Z}}_T^{(j)}\right) < -1$，否则 $D_T = 0$。根据步骤（1），$F > -1$ 在 $\mathcal{O} \supset \mathcal{K}_\delta$ 上，因此这要求 $\hat{\mathbf{Z}}_T^{(j)} \notin \mathcal{K}_\delta$；于是，由于 $\mathbf{z}_T^{(j)} \in \mathcal{K}$，$\left|\hat{\mathbf{Z}}_T^{(j)} - \mathbf{z}_T^{(j)}\right| > \delta$，对于某个 j。我们对该事件的概率关于 r 和 n一致地进行界定。$\hat{\mathbf{Z}}_T^{(j)} - \mathbf{z}_T^{(j)}$ 的每个元素都具有如下形式

$$\frac{4}{n^2}\sum_{\ell\in A'}\sum_{\ell'\in B'}\left(\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^{(\ell)}, \mathbf{v}^{(\ell')}\right) - \overline{\mathrm{Cov}}\left(\mathbf{v}^{(\ell)}, \mathbf{v}^{(\ell')}\right)\right), \qquad A', B' \in \{A_j, B_j\}, \tag{84}$$

并且，根据定理 1 证明中的 (4)，总和中的每个差值都可分解为 $Y_{\ell\ell'} + W_{\ell\ell'}$其中 $\mathbb{E}\left[Y_{\ell\ell'}^2\right] \le c_1'/T$和 $\mathbb{E}\left[|W_{\ell\ell'}|\right] \le c_2'/T$，其中 $c_1' := \left(4M\sqrt{C} + \sqrt{C^2 + K_4}\right)^2$ 以及 $c_2' := C$。将关于 $A' \times B'$ 的求和界定为对所有对 $(\ell, \ell') \in [n]^2$ 的求和，而后者与 j 无关，因此每个划分中的每个条目都满足

$$\max_j \left|\left[\hat{\mathbf{Z}}_T^{(j)} - \mathbf{z}_T^{(j)}\right]_q\right| \le \bar{Y} + \bar{W}, \text{ for } q = 1, 2, 3, \tag{85}$$

$$\bar{Y} := \frac{4}{n^2}\sum_{\ell,\ell'=1}^{n} |Y_{\ell\ell'}|, \qquad \bar{W} := \frac{4}{n^2}\sum_{\ell,\ell'=1}^{n} |W_{\ell\ell'}|, \tag{86}$$

其中，闵可夫斯基不等式给出 $\mathbb{E}\left[\bar{Y}^2\right] \le 16c_1'/T$。此外，$\mathbb{E}\left[\bar{W}\right] \le 4c_2'/T$。由于 $\hat{\mathbf{Z}}_T^{(j)} - \mathbf{z}_T^{(j)}$ 有三个条目，$D_T \ne 0$ 意味着 $\bar{Y} + \bar{W} > \frac{\delta}{\sqrt{3}}$（因为我们至少需要一个分量 q，满足 $\left|\left[\hat{\mathbf{Z}}_T^{(j)} - \mathbf{z}_T^{(j)}\right]_q\right| > \delta/\sqrt{3}$）。此外，$\bar{\bar{Y}} + \bar{W} > \delta/\sqrt{3}$

需要 $\bar{Y} > \frac{\delta}{2\sqrt{3}}$ 或 $\bar{W} > \frac{\delta}{2\sqrt{3}}$ 根据马尔可夫不等式，可得

$$\begin{aligned}\mathbb{P}\left(D_T \neq 0\right) &\leq \mathbb{P}\left(\bar{Y} > \frac{\delta}{2\sqrt{3}}\right) + \mathbb{P}\left(\bar{W} > \frac{\delta}{2\sqrt{3}}\right) \\ &\leq \frac{\mathbb{E}\left[\bar{Y}^2\right]}{\delta^2/12} + \frac{\mathbb{E}\left[\bar{W}\right]}{\delta/(2\sqrt{3})} \\ &\leq \frac{192c_1'}{\delta^2 T} + \frac{8\sqrt{3}c_2'}{\delta T} =: \frac{c_D}{T}. \end{aligned} \tag{87}$$

因此，由于 $|D_T| \leq 1$、$|\mathbb{E}\left[D_T\right]| \leq \mathbb{E}|D_T| \leq$ 上确界于 $|D_T|\,\mathbb{P}(D_T \neq 0)$ $\leq c_D/T$，类似地，$\mathbb{E}\left[D_T^2\right] \leq c_D/T$。

~~(ii) **对划分取平均**根据定理 1，将其应用于步骤 (2) 中的 F，对于每个划分，$\tilde{F}\left(\hat{\mathbf{Z}}_T^{(j)}\right) = F\left(\mathbf{z}_T^{(j)}\right) + \varepsilon_T^{(j)}$，其中 $\left|\mathbb{E}\left[\varepsilon_T^{(j)}\right]\right| \leq c/T$ 且 $\mathbb{E}\left[\left(\varepsilon_T^{(j)}\right)^2\right] \leq c/T$。由于 c 仅取决于 M、C、K_4 和 a_0，因此对所有划分都相同。取平均可得

$$\tilde{X} = \bar{x}_T + \bar{\varepsilon}_T, \qquad \bar{x}_T := \frac{1}{r}\sum_{j=1}^{r} F\left(\mathbf{z}_T^{(j)}\right), \qquad \bar{\varepsilon}_T := \frac{1}{r}\sum_{j=1}^{r} \varepsilon_T^{(j)}, \tag{88}$$

其中，$\bar{x}_T$ 是式（75）中平方根的被开方数。根据三角不等式 $|\mathbb{E}\left[\bar{\varepsilon}_T\right]| \leq \frac{1}{r}\sum_j \left|\mathbb{E}\left[\varepsilon_T^{(j)}\right]\right| \leq c/T$，以及闵可夫斯基不等式 $\mathbb{E}\left[\bar{\varepsilon}_T^2\right] \leq \left(\frac{1}{r}\sum_j \sqrt{\mathbb{E}\left[\left(\varepsilon_T^{(j)}\right)^2\right]}\right)^2 \leq c/T$.

（iii）平方根在一般情况下，由于 $\mathbb{E}\left[\left(\tilde{X} - \bar{x}_T\right)^2\right] = \mathbb{E}\left[\bar{\varepsilon}_T^2\right] \leq c/T$根据（ii），~~引理 3 与 $X_T = X$和 $x_T = \bar{x}_T$给出 g（X）$= g$（$\bar{x}_T$）$+\varepsilon_T^{\max}$，其中 $|\mathbb{E}\left[\varepsilon_T^{\max}\right]| \leq$（$c/T$）$^{1/4}$且 $\mathbb{E}\left[\right.$（$\varepsilon_T^{\max}$）$^2\left.\right] \leq$（c/T）$^{1/2}$。如果对所有 T都有 $\bar{x}_T \geq f_{\min}$，则改用引理 2，并取 $d = 1$。（ii）的展开式具有该引理所要求的形式，其中随机部分为 $Y_T :=\bar{\varepsilon}_T - \mathbb{E}\left[\bar{\varepsilon}_T\right]$，确定性部分为 $W_T :=\mathbb{E}\left[\bar{\varepsilon}_T\right]$：我们有 $\mathbb{E}\left[Y_T\right] = 0$、$\mathbb{E}\left[Y_T^2\right] \leq \mathbb{E}\left[\bar{\varepsilon}_T^2\right] \leq c/T$和 $\mathbb{E}\left[|W_T|\right] = |\mathbb{E}\left[\bar{\varepsilon}_T\right]| \leq c/T$。我们取 $\mathcal{K} = [f_{\min}, 1]$，它包含所有 $\bar{x}_T$，因为在总体对应量处（步骤（1））有 $0 \leq F \leq 1$，并且 $\delta = f_{\min}/2$。于是，g 在开集~~ ~~（$f_{\min}/4$，∞）$\supset \mathcal{K}_\delta$上无穷次可微，并且 $0 \leq g\left(X\right)$ g（$\bar{x}_T$）≤ 1，因为 $X \leq 1$。因此，引理 2 给出~~g（X）$= g$（$\bar{x}_T$）$+\varepsilon_T$。

最后，$\widehat{NC}^{\mathrm{SHSB}} = g(X) = g(\bar{x}_T) + \varepsilon_T^{\max} + D_T$（分别为 $g(\bar{x}_T) + \varepsilon_T + D_T$）。根据 (i)，$|\mathbb{E}\left[D_T\right]| \leq c_D/T$和 $\mathbb{E}\left[D_T^2\right] \leq c_D/T$。根据偏差的三角不等式和均方根误差的闵可夫斯基不等式，将 D_T加到 ε_T上会得到一个类型为 ε_T、常数为$(\sqrt{c} + \sqrt{c_D})^2$的误差项。将 D_T加到 $\varepsilon_T^{\max}$上会得到一个类型为 $\varepsilon_T^{\max}$、常数为 $\left(c^{1/4} + \max\left(c_D,\ \sqrt{c_D}\right)\right)^4$的误差项。

□

引理5（对所有划分取平均值）。令 $\mathcal{S}$ 为所有可能划分的集合。那么

$$\frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}} \Delta_{AB}^{w} = \frac{1}{n-1}\sum_k \widehat{\mathrm{Var}}_\ell\left[w_k\right]. \tag{89}$$

因此，式（73）分子 $\sum_k \bar{w}_k^2 - \Delta_{AB}^w$ 在所有划分上的平均值，恰好是 $\widehat{NC}^{\mathrm{ATV}}$ 的分子 $\sum_k \bar{w}_k^2 - \frac{1}{n-1}\sum_k \widehat{\mathrm{Var}}_\ell\left[w_k\right]$。

证明。对于一个划分（A, B），我们定义 $\epsilon_\ell^A := 1$ 如果 $\ell \in A$，否则为 -1。我们展开

$$\begin{aligned}
\frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\Delta_{AB}^{w} &= \frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\frac{1}{4}\sum_{k}\left(\bar{w}_k^A-\bar{w}_k^B\right)^2 \\
&= \frac{1}{4}\frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\sum_{k}\left(\frac{2}{n}\sum_{\ell}\epsilon_\ell^A w_k^{(\ell)}\right)^2 \\
&= \frac{1}{n^2}\frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\sum_{k}\left(\sum_{\ell}\left(w_k^{(\ell)}\right)^2+\sum_{\ell\neq\ell'}\epsilon_\ell^A\epsilon_{\ell'}^A w_k^{(\ell)}w_k^{(\ell')}\right) \\
&= \frac{1}{n^2}\sum_{k}\left(\sum_{\ell}\left(w_k^{(\ell)}\right)^2+\sum_{\ell\neq\ell'}w_k^{(\ell)}w_k^{(\ell')}\frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\epsilon_\ell^A\epsilon_{\ell'}^A\right). \qquad (90)
\end{aligned}$$

根据定义，当 ℓ 和 ℓ' 要么都在 A 中，要么都在 B 中时，$\epsilon_\ell^A\epsilon_{\ell'}^A$ 为 1；否则为 -1。据此，我们写作

$$\begin{aligned}
\frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\epsilon_\ell^A\epsilon_{\ell'}^A &= \frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\mathbb{1}_{\{\ell,\ell'\in A\text{ or }\ell,\ell'\in B\}}-\left(1-\frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\mathbb{1}_{\{\ell,\ell'\in A\text{ or }\ell,\ell'\in B\}}\right) \\
&= \frac{2}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\mathbb{1}_{\{\ell,\ell'\in A\text{ or }\ell,\ell'\in B\}}-1 = 4\cdot\frac{\binom{n-2}{n/2-2}}{\binom{n}{n/2}}-1 = 4\cdot\frac{n/2(n/2-1)}{n(n-1)}-1 \\
&= -\frac{1}{n-1}. \qquad (91)
\end{aligned}$$

将（91）代入（90），

$$\begin{aligned}
\frac{1}{|\mathcal{S}|}\sum_{(A,B)\in\mathcal{S}}\Delta_{AB}^{w} &= \frac{1}{n^2}\sum_{k}\left(\sum_{\ell}\left(w_k^{(\ell)}\right)^2-\frac{1}{n-1}\sum_{\ell\neq\ell'}w_k^{(\ell)}w_k^{(\ell')}\right) \\
&= \frac{1}{n^2}\sum_{k}\left(\frac{n}{n-1}\sum_{\ell}\left(w_k^{(\ell)}\right)^2-\frac{1}{n-1}\left(\sum_{\ell}w_k^{(\ell)}\right)^2\right) \\
&= \frac{1}{n-1}\sum_{k}\left(\frac{1}{n}\sum_{\ell}\left(w_k^{(\ell)}\right)^2-\left(\frac{1}{n}\sum_{\ell}w_k^{(\ell)}\right)^2\right) \\
&= \frac{1}{n-1}\sum_{k}\widehat{\mathrm{Var}}_\ell\left[w_k\right]. \qquad (92)
\end{aligned}$$

□

注3（与 ATV 的比较）。当 $\mathcal{S}$ 为所有划分的集合时，利用引理5,

$$\begin{aligned}
\sqrt{\frac{1}{r}\sum_{(A,B)\in\mathcal{S}}\frac{\sum_k\bar{w}_k^2-\Delta_{AB}^{w}}{\sum_k\bar{w}_k^2+\bar{\sigma}^2/n-\Delta_{AB}^{\mathrm{sd}}}} &\geq \sqrt{\frac{1}{r}\sum_{(A,B)\in\mathcal{S}}\frac{\sum_k\bar{w}_k^2-\Delta_{AB}^{w}}{\sum_k\bar{w}_k^2+\bar{\sigma}^2/n}} \\
&= \sqrt{\frac{\sum_k\bar{w}_k^2-\frac{1}{n-1}\sum_k\widehat{\mathrm{Var}}_\ell\left[w_k\right]}{\sum_k\bar{w}_k^2+\bar{\sigma}^2/n}}. \qquad (93)
\end{aligned}$$

这意味着，除误差项外，$\widehat{NC}^{\mathrm{SHSB}} \geq \widehat{NC}^{\mathrm{ATV}}$。当且仅当对于每个划分均满足 A, B 时，等式成立。

B.5 单试次噪声上限

在本节中，我们推导单试次响应 NC_ℓ^{single} 的真实噪声上限及其估计量 $\widehat{NC}_\ell^{\text{LOOC}}$ 和 $\widehat{NC}_\ell^{\text{TTMC}}$ 的闭式表达式。我们固定一个试次 ℓ，并假设 $n \geq 2$。记作

$$\mathbf{v}^{(-\ell)} := \mathbf{v}^{[n]\setminus\{\ell\}}, \qquad \mathbf{v}^{*(-\ell)} := \mathbf{v}^{*[n]\setminus\{\ell\}}, \qquad \bar{w}_k^{(-\ell)} := \bar{w}_k^{[n]\setminus\{\ell\}}, \qquad \bar{\sigma}^2_{(-\ell)} := \frac{1}{n-1}\sum_{\ell'\neq\ell}\sigma^2_{\ell'} \tag{94}$$

对于其他试次的平均值、其无噪信号对应项、其权重及其噪声水平。

B.5.1 真实噪声上限 NC^{single}

单试次响应的真实噪声上限，是无噪信号与同一试次记录信号之间的相关性，NC_ℓ^{single}：$= \widehat{\text{Corr}}_t\left(\mathbf{v}^{*(\ell)},\ \mathbf{v}^{(\ell)}\right)$。

命题4（真实噪声上限，单试次响应）。假设存在一个常数 $a_0 > 0$，使得 $\widehat{\text{Var}}_t\left(\mathbf{v}^{*(\ell)}\right) \geq a_0$ 对所有 T均成立。在假设 **1–4** 下，

$$NC_\ell^{\text{single}} = \sqrt{\frac{\widehat{\text{Var}}_t\left(\mathbf{v}^{*(\ell)}\right)}{\widehat{\text{Var}}_t\left(\mathbf{v}^{*(\ell)}\right) + \sigma_\ell^2}} + \varepsilon_T = \sqrt{\frac{\sum_k \left(w_k^{(\ell)}\right)^2}{\sum_k \left(w_k^{(\ell)}\right)^2 + \sigma_\ell^2}} + \varepsilon_T, \tag{95}$$

其中，误差项的常数 c 仅取决于 M、C、K_4 和 a_0。第二个表达式在定义 6 的表示下成立。与 NC^{mean}相比，首项不取决于 n：它仅由试次 ℓ 的信号和噪声决定。

证明。对于 $A = B = \{\ell\}$，我们应用推论 1 的 (45)，其中 $s(\{\ell\},\ \{\ell\}) = \sigma_\ell^2$，并按照命题 1 的证明进行化简。第二个表达式由引理4 和 $A = \{\ell\}$ 得出。 □

B.5.2 留一相关 $\widehat{NC}^{\text{LOOC}}$

LOOC 估计量将试次 ℓ 与其他试次的平均值 $\widehat{NC}_\ell^{\text{LOOC}}$ 进行相关：$= \widehat{\text{相关性}}_t\left(\mathbf{v}^{(\ell)},\ \mathbf{v}^{(-\ell)}\right)$。

命题5 (LOOC)。假设存在一个常数 $a_0 > 0$，使得对于所有 T，$\widehat{\text{Var}}_t\left(\mathbf{v}^{*(\ell)}\right) \geq a_0$ 以及 $\widehat{\text{Var}}_t\left(\mathbf{v}^{*(-\ell)}\right) \geq a_0$。在假设 **1–4** 下，

$$\widehat{NC}_\ell^{\text{LOOC}} = \frac{\sum_k w_k^{(\ell)}\bar{w}_k^{(-\ell)}}{\sqrt{\left(\sum_k \left(w_k^{(\ell)}\right)^2 + \sigma_\ell^2\right)\left(\sum_k \left(\bar{w}_k^{(-\ell)}\right)^2 + \frac{\bar{\sigma}^2_{(-\ell)}}{n-1}\right)}} + \varepsilon_T \tag{96}$$

其中，误差项的常数 c 仅取决于 M、C、K_4 和 a_0。

证明。我们将推论 1 中的 (44) 应用于 $A = \{\ell\}$ 和 $B = [n] \setminus \{\ell\}$。这两个集合不相交，因此 $s(A, B) = 0$，而 $s(A,\ A) = \sigma_\ell^2$，并且

$$s(B,B) = \frac{1}{(n-1)^2}\sum_{\ell'\neq\ell}\sigma^2_{\ell'} = \frac{\bar{\sigma}^2_{(-\ell)}}{n-1}. \tag{97}$$

引理4给出 $\widehat{\text{Cov}}_{t_2}\left(\mathbf{v}^{*(\ell)},\ \mathbf{v}^{*(-\ell)}\right) = \sum_k w_k^{(\ell)}\bar{w}_k^{(-\ell)}$、$\widehat{\text{Var}}_t\left(\mathbf{v}^{*(\ell)}\right) = \sum_k \left(w_k^{(\ell)}\right)^2$ 以及 $\widehat{\text{Var}}_t\left(\mathbf{v}^{*(-\ell)}\right) = \sum_k \left(\bar{w}_k^{(-\ell)}\right)^2$。 □

注释 4。在 n 较大的极限下，$\widehat{NC}_\ell^{\text{LOOC}}$ 趋近于

$$NC_\ell^{\text{single}} \cdot \cos_k\left(w^{(\ell)}, \overline{w}^{(-\ell)}\right) \leq NC_\ell^{\text{single}}, \tag{98}$$

误差项 ε_T，并且

$$\cos_k(x,y) := \frac{\sum_k x_k y_k}{\sqrt{\sum_k x_k^2 \sum_k y_k^2}} \tag{99}$$

k之间的余弦相似度。等式成立当且仅当对于某个 $\lambda > 0$， $w^{(\ell)} = \lambda\, w^{(-\ell)}$，而这通常仅在同质状态下满足。

B.5.3 试次对均值相关 $\widehat{NC}^{\mathrm{TTMC}}$

TTMC 估计量将试次 ℓ 与包括试次 ℓ 本身在内的所有试次的平均值进行相关分析，$\widehat{NC}_\ell^{\mathrm{TTMC}} := \widehat{\mathrm{Corr}}_t\left(\mathbf{v}^{(\ell)},\ \bar{\mathbf{v}}\right)$。

命题6（TTMC）。假设存在一个常数 $a_0 > 0$，使得对于所有 T， $\widehat{\mathrm{Var}}_t\left(\mathbf{v}^{*(\ell)}\right) \geq a_0$ 以及 $\widehat{\mathrm{Var}}_t\left(\bar{\mathbf{v}}^*\right) \geq a_0$。在假设 1–4 下，

$$\widehat{NC}_\ell^{\mathrm{TTMC}} = \frac{\sum_k w_k^{(\ell)} \bar{w}_k + \frac{\sigma_\ell^2}{n}}{\sqrt{\left(\sum_k \left(w_k^{(\ell)}\right)^2 + \sigma_\ell^2\right)\left(\sum_k \bar{w}_k^2 + \frac{\bar{\sigma}^2}{n}\right)}} + \varepsilon_T \tag{100}$$

其中，误差项的常数 c 仅取决于 M、 C、 K_4 和 a_0。

证明。我们将推论 1 中的（44）式应用于 $A = \{\ell\}$ 和 $B = [n]$。现在 $A \cap B = \{\ell\}$，因此

$$s(A,B) = \frac{\sigma_\ell^2}{n}, \qquad s(A,A) = \sigma_\ell^2, \qquad s(B,B) = \frac{1}{n^2}\sum_{\ell'=1}^{n} \sigma_{\ell'}^2 = \frac{\bar{\sigma}^2}{n}. \tag{101}$$

引理4给出 $\widehat{\mathrm{Cov}}_t\left(\mathbf{v}^{*(\ell)},\ \bar{\mathbf{v}}^*\right) = \sum_k w_k^{(\ell)} \bar{w}_k$。 □

注5。当 n较大时的极限下， $\widehat{NC}_\ell^{\mathrm{TTMC}}$ 趋近于

$$NC_\ell^{\mathrm{single}} \cdot \cos_k\left(w^{(\ell)}, \overline{w}\right) \leq NC_\ell^{\mathrm{single}}, \tag{102}$$

在误差项 ε_T范围内。因此，与 $\widehat{NC}_\ell^{\mathrm{LOOC}}$ 类似， $\widehat{NC}_\ell^{\mathrm{TTMC}}$ 仅在同质状态下收敛于真实噪声上限。

注6。在异质状态下， $\widehat{NC}_\ell^{\mathrm{TTMC}}$ 显然不是噪声上限的上界，因为它收敛于一个低于 $NC_\ell^{\mathrm{single}}$ 的极限。此外，即使在同质情况下，

$$\widehat{NC}_\ell^{\mathrm{TTMC}} = NC_\ell^{\mathrm{single}} \cdot \underbrace{\frac{1 + \frac{\sigma_\ell^2}{n\sum_k w_k^2}}{\sqrt{1 + \frac{\bar{\sigma}^2}{n\sum_k w_k^2}}}}_{\gamma} + \varepsilon_T, \tag{103}$$

其中 γ 在 $2\sigma_\ell^2 < \sigma^2$ 时可能小于 1。

B.5.4 留一预测 $\widehat{NC}_\ell^{\mathrm{LOOP}}$

由于 LOOP 涉及拟合回归，在我们的框架中无法得到一般的闭式形式。我们在此定性地描述其渐近行为。与其他估计量不同，LOOP 根据其他试次中的所有单元来预测一个单元。因此，在本节中，我们保留单元索引$i = 1, \ldots\ldots m$。我们假设 B.2 节中的假设对每个单元都成立，同时允许同一试次中不同单元的噪声存在相关性。与 B.4.2 节类似，我们假设每条记录都已在时间维度上进行中心化。

对于目标单元 i 和试次 ℓ， $p := m(n-1)$ 预测变量是所有单元$j \in [m]$ 的所有其他试次 $\ell' \neq \ell$的记录数据 $\mathbf{v}^{(\ell')\,j}$。在一组训练时间点上，通过普通最小二乘拟合 LOOP(OLS) 或岭回归拟合带截距的线性模型，并且 $NC_{i,\ell}$ 是该模型的预测与 $\mathbf{v}_i^{(\ell)}$ 在一组测试时间点上的相关性。由于同一试次中不同单元的噪声可能存在相关性，我们将假设4扩展到单元对：对于所有单元 j,k 和试次 ℓ，$\sup_t \sum_{s=1}^{\infty} \left|\mathrm{Cov}\left(\eta_j^{(\ell)}{}_{,t}\, \eta_k^{(\ell)}{}_{t,},\ \eta_j^{(\ell)}{}_{,s}\, \eta_k^{(\ell)}{}_{,s}\right)\right| < K_4$。

时间点数量较大 设 T和 T' 分别为训练时间点和测试时间点的数量，$X \in \mathbb{R}^{T\times p}$、$X' \in \mathbb{R}^{T'\times p}$为相应的预测变量矩阵。我们将 $\chi^{(-\ell)} \subset \mathbb{R}^{T'}$记作由 X'的列张成的空间，并将 $P^{(-\ell)}$ 记作向 $\chi^{(-\ell)}$的正交投影。我们考虑极限 $T, T' \to \infty$，其中 $p/T \to 0$和 $p/T' \to 0$。假设训练点和测试点是随机选取的，则在此极限下，在 T或 T' 上计算的$\sigma^2_{i,\ell}$和 $\widehat{\text{Var}}_t(\mathbf{v}_i^{*(\ell)})$是等价的。因此，预测值 $\hat{\mathbf{v}}_i$由目标在 $\chi^{(-\ell)}$上的正交投影 $P^{(-\ell)}\mathbf{v}_i^{(\ell)}$给出。由于目标的噪声 $\boldsymbol{\eta}_i^{(\ell)}$与预测变量相互独立，目标的投影渐近等于$P^{(-\ell)}\mathbf{v}_i^{*(\ell)}$。估计量 $\widehat{NC}^{\text{LOOP}}_{i\,\ell}$,随后由下式给出

$$\widehat{\text{Corr}}_t\left(\mathbf{v}_i^{(\ell)}, \hat{\mathbf{v}}_i^{(\ell)}\right) = \sqrt{\frac{\widehat{\text{Var}}_t\left(P^{(-\ell)}\mathbf{v}_i^{*(\ell)}\right)}{\widehat{\text{Var}}_t\left(\mathbf{v}_i^{*(\ell)}\right) + \sigma^2_{i,\ell}}} = \frac{\left|P^{(-\ell)}\mathbf{v}_i^{*(\ell)}\right|}{\left|\mathbf{v}_i^{*(\ell)}\right|} \cdot NC^{\text{single}}_{i,\ell} = \cos\left(\alpha^{(\ell)}\right) \cdot NC^{\text{single}}_{i,\ell}, \tag{104}$$

其中，$|.|$ 是欧几里得范数，$\alpha^{(\ell)}$ 是$\mathbf{v}_i^{*(\ell)}$ 与 $\chi^{(-\ell)}$之间的夹角，第二个等式在$NC^{\text{single}}_{i\,\ell}$的误差项范围内成立。因此，$\widehat{NC}^{\text{LOOP}}_{i\,\ell}$是单试次噪声上限的下界，当且仅当无噪声目标位于预测变量的张成空间中时，该下界才能达到。两个效应会使 $\cos\left(\alpha^{(\ell)}\right) < 1$。首先，预测变量带有噪声，因此 $\chi^{(-\ell)}$ 由可重复响应的含噪版本张成，而不是由可重复响应本身张成。其次，在异质状态下，$\mathbf{v}_i^{*(\ell)}$ 很可能不包含在可重复响应的张成空间中。特别是，如果 $K > p$潜在信号 $\boldsymbol{\phi}_1$、……、 $\boldsymbol{\phi}_K$无法从 $\{\mathbf{v}_i^{*\ell'}\}_{\ell'\neq\ell、\ i=1……、\ m}$中恢复，那么可能无法从可重复信号中获得$\mathbf{v}_i^{*(\ell)} = \sum_k w_{ik}^{(\ell)}\boldsymbol{\phi}_k$。在此极限下，$\widehat{NC}^{\text{LOOP}}_{i\,\ell}$ 至少与 $\widehat{NC}^{\text{LOOC}}_{i\,\ell}$ 一样大。

大量试次 随着 n 的增加，在 $T/p \to \infty$的情况下，这两种效应都可能减弱：预测变量包含越来越多的相同潜在信号的噪声副本，从而可以将其中的噪声平均掉；而其他试次的可重复响应则张成不断扩大的子空间。 $\alpha^{(\ell)}$趋近于$\mathbf{v}_i^{*(\ell)}$ 与可重复响应 $\{\mathbf{v}_i^{*\ell'}\}_{\ell'\neq\ell、\ i=1……m}$的张成空间之间的夹角；当$\mathbf{v}_i^{*(\ell)}$ 位于该空间中时，$\widehat{NC}^{\text{LOOP}}_{i\,\ell}$、可以恢复真实噪声上限。与 $\widehat{NC}^{\text{LOOC}}_{i\,\ell}$、和 $\widehat{NC}^{\text{TTMC}}_{i\,\ell}$、不同，$\widehat{NC}^{\text{LOOP}}_{i\,\ell}$、向 $NC^{\text{single}}_{i\,\ell}$、的收敛并不局限于同质状态。

有限数量的时间点 对于有限的 T，泛化误差会降低测试集上的相关性。当单元的噪声很大时，这种泛化误差也更大，并且会随预测变量的数量$p = m(n-1)$增加而增大。在固定 T 的情况下，随着试次数 n增加，$\widehat{NC}^{\text{LOOP}}_{i\,\ell}$、因此同时包含一个因 $\alpha^{(\ell)}$ 衰减而产生的递增项，以及一个递减的泛化项。当 p/T不够小时，它可能低于没有拟合参数的 $\widehat{NC}^{\text{LOOC}}_{i\,\ell}$、。岭回归会降低泛化误差，但会以偏差取而代之。

外推至 $n \to \infty$ 这些结果解释了为什么 LOOP 对 n 的依赖不存在简单的通用形式。极限项通过公式 (104) 中的夹角取决于试次ℓ 的潜在信号如何由其他试次表示。此外，在 T固定时，泛化误差会随 n增大。拟合 $\widehat{NC}^{\infty}_{i\ ,\ \ell}\left(1 - e^{-n/\tau_0}\right)$ 假设饱和速度非常快，并忽略了泛化项，这两点都会导致对噪声上限的低估。

C 合成数据参数

表 4 给出了生成图 2 所用合成数据集的全部参数。这些参数在第 3.4 节中定义。

D 经验有限-T 收敛

本节在Sherlock数据集上对噪声上限估计量随刺激长度 T 增长的收敛速率进行实证评估。图 3 展示了 SHSB、ATV、LOOC 和 TTMC 随刺激长度 T 的变化。如果某个估计量由一个与 T 无关的主导项和

表 4：每个数据集的模拟参数。

数据集	T	n	ρ_{shared}	σ_0	s	σ_w	Δ	信号
THINGS fMRI1	100	12	0.30	4.00	1.0	1.00	0	高斯
THINGS EEG2	200	80	0.15	2.75	0.9	1.00	0	高斯
Sherlock V1	1976	16	0.60	0.60	0.6	2.00	5	类BOLD
Sherlock IT	1976	16	0.60	1.90	1.2	0.85	5	类BOLD

$\Delta = 0$：同质状态；$\Delta > 0$：异质状态。

一个期望阶为 $1/T$ 的误差组成，那么将 $a + bT^{-p}$ 拟合到其平均值上，应当在 T 足够大时得到 $p \approx 1$。对于 V1区和 IT区中的 LOOC 与 TTMC，我们确实观察到了这一点，且从 $T < 100$ 开始便已如此。SHSB 和 ATV 的表现有所不同：在 V1区，p 接近 1，但在 IT区则更低。这就是第 3.2 节所描述的瞬态区域。对于信号较弱的单元，平方根的阈值首先会产生阶为 $T^{-1/4}$ 的正偏差，只有当有限-T 波动相对于信号足够小时，才会达到 $1/T$ 速率。在 V1区，由于信号较强，所示 T 范围内的大多数体素已经处于 $1/T$区域。在 IT区，由于信号较弱，体素平均值混合了具有阶为 $1/T$ 的负偏差的体素和具有阶为 $T^{-1/4}$ 的正偏差的体素。结果是负偏差，其衰减速度慢于 $1/T$。

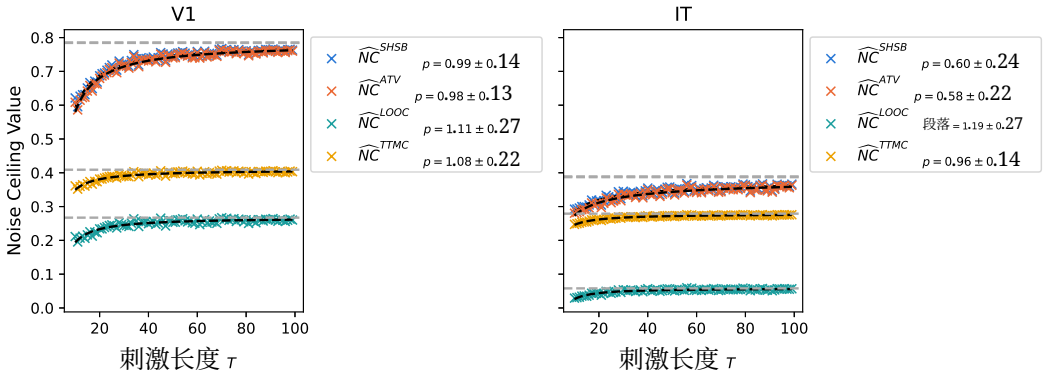

图3：**估计量随刺激长度的收敛情况 T（Sherlock数据集）**。每个十字表示在长度为 T 的100个时间窗口上计算某个估计量，并在体素上取平均所得的平均值。每个 T均使用全部16名被试，因此只有刺激长度 T 发生变化。$\widehat{NC}^{\text{LOOC}}$ 和 $\widehat{NC}^{\text{TTMC}}$ 在每名留出受试者上取平均。对于每个估计量，均值$\widehat{\ }$曲线使用 $NC(T) = a + bT^{-p}$（黑色虚线）进行拟合，其中 a 是在较大 T 时达到的渐近值（灰色虚线），p 为收敛速率。图例给出 p 及其95%置信区间（± 1.96 标准误，来自最小二乘拟合的协方差）。

E 噪声上限计算检查清单

这里列出了计算噪声上限时常见的陷阱，以及如何避免这些问题。

SHSB 和 ATV

- **平方根：** ATV 和 SHSB 都包含一个平方根。当脑分数计算为皮尔逊相关时，这个平方根是必需的，但它经常被省略（van Bree 等，2025）。当脑分数计算为 anR^2 时，不应应用平方根。

ATV

- **中心化：** 在计算跨试次的方差之前，必须对每个试次的记录数据进行时间中心化。

- **归一化：** 试验间方差必须采用无偏归一化 $1/(n-1)$ 进行计算。使用 $1/n$时，估计量会产生向上的偏差。

SHSB

- **对划分取平均值：** 要使 SHSB 能够较好地估计噪声上限，就需要对许多划分取平均值。为避免引入额外偏差，应在进行 $\frac{2\rho}{1+\rho}$ 校正之后、阈值处理和平方根运算之前执行此平均操作。

- **试次数量为奇数：** 对于 n 奇数情况，应优先选择 ATV，而不是 SHSB。

TTMC

- **错误的上界：** 当试次对应于不同被试时，不应使用 TTMC，尤其不应将其视为噪声上限的上界。

LOOP

- **训练集与测试集的独立性：** 对于自相关记录，随机选取的训练时间点和测试时间点之间存在相关性。测试集应改为由连续时间点块组成，并通过未使用的间隔与训练集分隔开；这些间隔应长于信号和噪声的自相关时长。拟合编码模型时同样适用。

- **预处理：** 某些预处理步骤（例如高通滤波器）会将信息扩散到多个时间点，因此可能在训练集和测试集之间引入相关性。应分别对每个数据集应用这些步骤。

- **在估计量中进行选择：** 对于每个单元，选取 LOOC、LOOP 及其变体中的最高值时，必须在不同于测试集的验证集上进行选择，以避免使估计值产生向上的偏差。

报告脑评分

- **下界：** 当只能获得噪声上限的一个下界时，归一化脑得分应报告为一个区间，即从未归一化得分到按下界归一化后的得分，并将上限截断为 1。