arXiv:2611.28865v1 [stat.CO] 21 Nov 2026

多变量感觉辨别模型的快速似然评估

Philip Greengard[1], Fangfang Hong[2], David H. Brainard[2], 和 Alex Williams[3,1]

[1]计算神经科学中心，Flatiron Institute[2]心理学系，宾夕法尼亚大学[3]神经科学中心，纽约大学

2026年11月24日

摘要

感觉辨别通常通过拟合带有似然函数的统计模型来衡量，而这些似然函数在计算上往往难以评估。在标准的三角测试和二三组测试中，会向受试者呈现三个刺激，其中两个属于同一种类，另一个属于另一种类，并要求受试者找出其中不同的一个（三角测试），或者判断两个刺激中哪一个与参照刺激相匹配（二-三点检验）。在瑟斯顿模型中，受试者对每个刺激的内部表征服从高斯分布，其均值和协方差取决于刺激；受试者作答正确的概率是对一个 $3d$ 维高斯分布进行积分，而对于 $d > 1$ 没有闭式解，其中 d 是刺激空间的维数。为了将统计模型拟合到从三角测试或二-三点检验中收集的数据，似然函数以及通常还包括其梯度都需要多次评估，这使统计推断效率低下。本文提出了用于评估三角测试和二-三点检验中某些类别似然函数的快速算法。这些算法依赖于降低这些函数的维数。三角形概率是 $2d + d\,(d + 1)$ 个参数的函数，可通过解析方法化为 $2d$ 个参数的函数；当两个刺激具有相同的协方差时，二选一概率对于任意 d 都可化为单个标量的函数。在降低这些函数的维数后，我们使用神经网络和切比雪夫展开来高效评估低维函数。在准确度相当的情况下，无论是进行单次评估，还是用于计算颜色辨别模型的最大后验（MAP）估计，我们的算法相较于现有方法都实现了数量级的加速。我们所采用的降维方法可推广到具有椭圆等值集的非高斯噪声模型，以及包含三个以上刺激的测试。

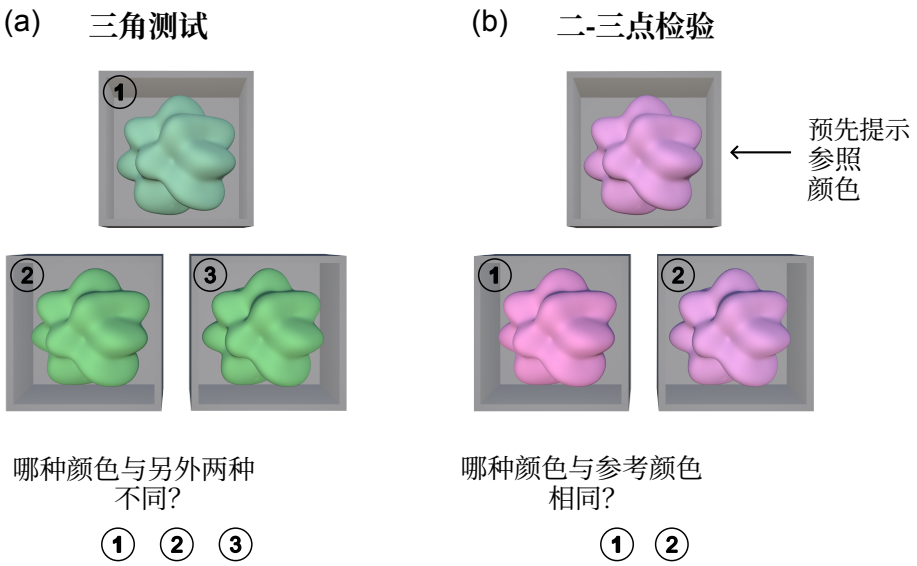


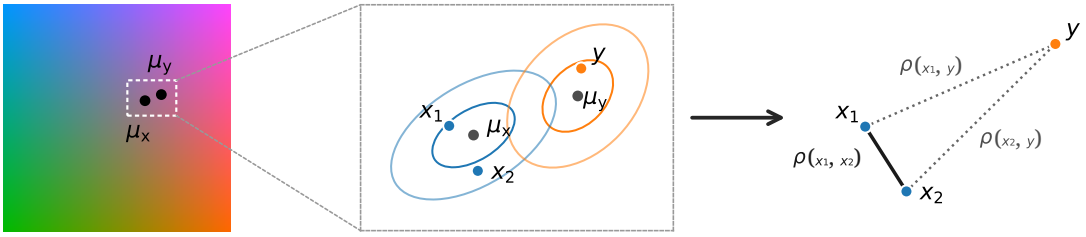
图 1：颜色辨别实验示例：(a) [22]的三角测试和 (b) [35]的二-三点检验。

1 引言

感觉辨别实验有着悠久的历史，例如应用于颜色感知 [31], 味觉 [15], 以及重量感知 [14]。在感觉辨别实验中，三角测试和二-三点检验是衡量两个刺激之间感知差异的标准选择 [38]。在三角测试中，会向受试者呈现三个刺激，其中两个相同，第三个不同。受试者需要找出与另外两个不同的刺激。在二-三点检验中，首先向受试者呈现一个参照刺激。随后再呈现两个刺激，其中一个与参照刺激相同，另一个不同。受试者需要判断两个刺激中哪一个与参照刺激相同。这些测试广泛应用于各种工业和学术研究场景。例如， [22] 的颜色辨别任务就是一种三角测试。在公共卫生研究中，人们会开展类似的感觉辨别任务，以识别味觉 [40, 34, 29, 20] 和嗅觉 [24, 23]方面的感知差异。在工业环境中，三角测试和二-三点检验经常用于质量控制 [26, 25]。

一般而言，刺激空间是 $\mathcal{X} \subseteq \mathbb{R}^d$ ，适用于某些 d，例如颜色任务，其中$d \leq 3$ [22]。研究人员将三角测试和二-三点检验等测试与统计模型结合起来，用于估计感官知觉辨别如何随刺激空间而变化。最常见的方法可追溯至瑟斯通 [43] ，也是信号检测理论的基础；[19, 32], 模型将每次刺激呈现视为产生一个从高斯分布中抽取的含噪内部表征。该模型的多变量推广已得到充分确立 [9, 36, 2], ，并已应用于多个领域 [1, 16,42, 22]。然而，在三元组任务的多变量瑟斯通模型下计算选择概率并非易事，且计算量很大，这很可能限制了这些模型的应用。

在三元　角度测试中，三个呈现的刺激里有两个相同，我们将其记为



(a) $\mathcal{X}$中的刺激　　(b) 内部表征　　(c) 决策规则

图 2：三角测试中瑟斯顿模型和决策规则的示意图。(a) 刺激空间 $\mathcal{X}$ 中的重复刺激 μ_{x} 和独特刺激 μ_{y} 。(b) 受试者对刺激的内部表征，x_1、$x_2 \sim \mathcal{N}(\mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})$ 和 $y \sim \mathcal{N}(\mu_{\mathrm{y}}, \Sigma_{\mathrm{y}})$。椭圆是高斯概率密度函数的等值线。(c) 决策规则：受试者在 ρ 下选择不属于最近点对的抽签结果，此处为 y。

x_1，$x_2 \sim \mathcal{N}(\mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})$，受试者对重复刺激的相互独立的内部表征。我们将独立于 x_1，x_2 的 $y \sim \mathcal{N}(\mu_{\mathrm{y}}, \Sigma_{\mathrm{y}})$ 记为独特刺激的内部表征。受试者在三角测试中的回答，即其对哪个是独特刺激的猜测，被建模为内部表征与另外两个刺激的内部表征距离最远的刺激，其中距离通过某个距离函数 ρ 来衡量：$\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$。也就是说，我们将三角测试中正确回答的概率记为 p_{tri}，其为

$$p_{\mathrm{tri}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma_{\mathrm{x}}, \Sigma_{\mathrm{y}}) = \Pr\big(\rho(x_1, x_2) < \min\{\rho(x_1, y),\, \rho(x_2, y)\}\big). \tag{1}$$

图 2 展示了一个试次所对应的模型。

在二-三点检验中，受试者指出所呈现的两个刺激中哪一个更接近参照刺激。也就是说，我们将 $x_{\mathrm{ref}} \sim \mathcal{N}(\mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})$ 定义为受试者对参照刺激的内部表征，将 $x_1 \sim \mathcal{N}(\mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})$ 定义为参照刺激的一个重复样本，将 $y_1 \sim \mathcal{N}(\mu_{\mathrm{y}}, \Sigma_{\mathrm{y}})$ 定义为另一个刺激，且三者相互独立。于是，p_{duo}（二-三点检验中正确回答的概率）为

$$p_{\mathrm{duo}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma_{\mathrm{x}}, \Sigma_{\mathrm{y}}) = \Pr\left(\rho(x_{\mathrm{ref}}, x_1) < \rho(x_{\mathrm{ref}}, y_1)\right). \tag{2}$$

模型中的一个重要考量是选择式（1）、（2）的距离函数 ρ。[36, 28, 38] 模型选择 ρ 为欧几里得距离，这相当于假设刺激空间经过参数化，使得知觉距离为欧几里得距离。对于$d \geq 2$，这种参数化仅在度量 $\Sigma(x)^{-1}$ 具有零黎曼曲率的情况下存在，其中 $\Sigma(x)$是刺激 x 的内部表征的协方差。即使在曲率确实为零的情况下，使ρ 成为欧几里得的参数化也由协方差场 $\Sigma(x)$决定，而这正是需要估计的量，因此实验者无法预先获得。事实上，在呈现刺激所采用的坐标中，辨别能力通常会随参照刺激 [31, 22],而变化 ρ，这表明在该表征中，欧几里得 ρ 并不是合适的距离函数。

[22, 35] 的颜色辨别模型使用形式为 ρ 的

$$\rho(x, y) = (x - y)^\top S^{-1}(x - y), \tag{3}$$

其中， S 是 Σ_{x} 和 Σ_{y}的凸组合。选择 ρ 是很自然的，因为 (i) 呈现给受试者的刺激 μ_{x}，μ_{y} 足够相近，以至于 Σ_{x}， Σ_{y} 近似相等；(ii) 对于任意 Σ_{x}， Σ_{y}，理想观察者 [19, 17]（按附录 A 中所述的意义）会通过 $(u-v)^{\top}\Sigma_{\mathrm{x}}^{-1}(u-v)$ 比较内部表征 u， v。

在许多采用三角检验和二-三点检验的实验中，分析实验数据时的主要计算挑战之一，是准确且高效地计算伯努利概率（1）和（2）。这些函数是在 $\mathbb{R}^{3d}$的不同区域上对同一个被积函数进行积分。具体而言，

$$p_{\mathrm{tri}}(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}})=\int_{S_{\mathrm{tri}}}\phi(u_1;\mu_{\mathrm{x}},\Sigma_{\mathrm{x}})\,\phi(u_2;\mu_{\mathrm{x}},\Sigma_{\mathrm{x}})\,\phi(v;\mu_{\mathrm{y}},\Sigma_{\mathrm{y}})\,\mathrm{d}u_1\,\mathrm{d}u_2\,\mathrm{d}v, \tag{4}$$

$$p_{\mathrm{duo}}(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}})=\int_{S_{\mathrm{duo}}}\phi(u_1;\mu_{\mathrm{x}},\Sigma_{\mathrm{x}})\,\phi(u_2;\mu_{\mathrm{x}},\Sigma_{\mathrm{x}})\,\phi(v;\mu_{\mathrm{y}},\Sigma_{\mathrm{y}})\,\mathrm{d}u_1\,\mathrm{d}u_2\,\mathrm{d}v, \tag{5}$$

其中， $\phi(u;\mu,\ \Sigma)$表示均值为 μ、协方差为 Σ 的高斯概率密度函数（p.d.f.），并且

$$S_{\mathrm{tri}}=\big\{(u_1,u_2,v)\in\mathbb{R}^{3d}\ :\ \rho(u_1,u_2)<\min\{\rho(u_1,v),\,\rho(u_2,v)\}\big\}, \tag{6}$$

$$S_{\mathrm{duo}}=\big\{(u_1,u_2,v)\in\mathbb{R}^{3d}\ :\ \rho(u_1,u_2)<\rho(u_1,v)\big\}. \tag{7}$$

关于在 ρ 为欧几里得距离 [9, 5, 6]的情况下准确且高效地计算函数（1）、（2）的算法，已有大量研究。在一维情况下（$d=1$），当假设所有 $x\in\mathcal{X}$的方差均为常数时，函数（1）、（2）可以高效计算——（1）是被积函数光滑的一维积分，而（2）是误差函数 [5, 6]的二阶多项式。对于 $d>1$，概率p_{tri}， p_{duo} 没有闭式表达式，且每次计算似然函数时通常都需要计算成本高昂的数值方法。在 [12, 28] 中针对 p_{tri}、在 [38]中针对 p_{tri}， p_{duo} 均介绍了求积方法。对于二选一概率， [10] 利用若干数学和数值方面的观察，将积分化为一个能够准确且高效计算的级数。蒙特卡洛方法是计算 p_{tri} 和 p_{duo}[11, 12, 22, 35] 的另一种方法，其标准误为 $\mathcal{O}\left(n^{-1/2}\right)$，使用 n个样本。

此处，我们介绍用于评估 p_{tri}， p_{duo} 的快速算法，其中 $d>1$ 且 ρ 具有形式（3）。这些算法所依据的主要数学观察是， p_{tri} 和p_{duo} 在刺激空间的仿射变换下保持不变。也就是说，对于任意可逆的 A 和任意 $b\in\mathbb{R}^d$，我们有

$$p_{\mathrm{tri}}(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}})=p_{\mathrm{tri}}(A\mu_{\mathrm{x}}+b,A\mu_{\mathrm{y}}+b,A\Sigma_{\mathrm{x}}A^{\top},A\Sigma_{\mathrm{y}}A^{\top}), \tag{8}$$

$$p_{\mathrm{duo}}(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}})=p_{\mathrm{duo}}(A\mu_{\mathrm{x}}+b,A\mu_{\mathrm{y}}+b,A\Sigma_{\mathrm{x}}A^{\top},A\Sigma_{\mathrm{y}}A^{\top}). \tag{9}$$

因此， p_{tri} 和 p_{duo}是 $2d+d(d+1)$ 个参数的函数，可以化简为 $2d$ 个变量的函数（见命题 1）。具体而言，通过适当选择可逆的 A，可以消去 d^2 个变量；再通过平移均值消去另一个 d。[1]

[1]我们注意到，在假设欧几里得 ρ的情况下，对于任意协方差 Σ_{x}， Σ_{y}（如 [10]所示），无法实现这种 p_{tri}， p_{duo} 维数的约简。对于欧几里得 ρ，函数 p_{tri}， p_{duo}在变换（8）下保持不变，其中 A 是正交矩阵的标量倍数。通过使用适当的 A，我们可以根据 $d\left(d-1\right)/2+1$ 个参数，将 p_{tri}， p_{duo} 约简为正交 $d\times d$矩阵的标量倍数的维数。平移并通过 A进行变换后，我们得到一个具有 $2d+d\left(d+1\right)/2-1$ 个参数的函数。

经过此变换后，我们使用神经网络来近似变换后的 p_{tri}（它是 $2d$ 个参数的函数）；针对 $d=2$，神经网络包含 369 个参数，针对 $d=3$，则包含 401 个参数。我们使用通过蒙特卡洛方法（ $n=10^5$ 个样本）标注的 3×10^5 个点训练每个网络。对于 $d=2$，神经网络在 $0.04\,\mu$ 秒内计算 p_{tri}，均方根误差小于 10^{-3}（见第 3.2 节）。相比之下，蒙特卡洛方法使用 10^5 个样本，达到相近的误差大约需要 1,300 μ 秒。我们实现的算法使用式（3）的 ρ 和 $S=\frac{2}{3}\Sigma_{\mathrm{x}}+\frac{1}{3}\Sigma_{\mathrm{y}}$，与 [22], 中的做法一致，不过所有分析对于 $\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}}$ 的任意凸组合都成立。我们使用该算法计算 p_{tri}，以求得 [22] 颜色辨别模型的最大后验估计。我们的算法在 0.30 秒内得到最大后验估计，其协方差场相对误差为 0.8%；相比之下，蒙特卡洛方法耗时 44 分钟，相对误差为 2.9%（第3节）。

对于具有 $\Sigma_{\mathrm{x}}=\Sigma_{\mathrm{y}}$ 的二-三点检验，函数 p_{duo} 仅通过标量 $r=(\mu_{\mathrm{y}}-\mu_{\mathrm{x}})^{\top}\Sigma_{\mathrm{x}}^{-1}(\mu_{\mathrm{y}}-\mu_{\mathrm{x}})$ 依赖于刺激和协方差。变量变换后，我们用 r 中的切比雪夫展开近似 p_{duo}，其阶数由用户指定的精度 ϵ 决定。切比雪夫展开的系数通过自适应求积算法计算：该算法利用变量变换后得到的一维积分来求值 p_{duo}。对于 $d=3$ 和 $\epsilon=10^{-8}$，包含 36 项的展开可在 $0.028\,\mu$ 秒内以小于 10^{-11} 的误差计算 p_{duo}（第 4 节）。

我们考虑的两种检验——三角测试和二-三点检验——以及假设协方差相等的选项 $\Sigma_{\mathrm{x}}=\Sigma_{\mathrm{y}}$，代表了感觉辨别模型中常见的四种情形。我们讨论其中三种：协方差相等或不等时的三角测试（第3节），以及协方差相等时的二-三点检验（第 4 节）。对于第四种情形，即协方差不等时的二-三点检验，已有相关分析和高效算法 [37, 10]。虽然本文不讨论这种情形，但针对 p_{tri} 和 p_{duo} 所提出的方法也适用于该情形。

本文其余部分的组织如下。第2节介绍贯穿全文其余部分所使用的理论工具，包括 p_{tri}，p_{duo}的仿射不变性。第3节和第 4 节介绍用于高效评估 p_{tri}，p_{duo}的快速算法。在第5节中，我们讨论扩展和局限性。最后，附录 A 将距离函数的选择与“理想观察者”的选择联系起来，并对“理想观察者”这一概念进行了严格界定。

本文的算法——第 3.2 节的神经网络和第 4 节的切比雪夫展开——均使用 Python 实现，并公开发布于 https: //github.com/pgree/fast-likelihood-sensory-discrimination。该代码仓库还包含用于生成本文图表的代码。

2 数学工具

本节的主要目的是证明命题 1，以降低 p_{tri} 和 p_{duo} 的参数维数。首先，我们将介绍若干辅助性的解析结果，这些结果将在命题 1 的证明中用到。在本节中，我们用 $\phi(x;\,\mu,\,\Sigma)$ 表示高斯概率密度函数，

$$\phi(x;\mu,\Sigma)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}(x-\mu)^{\top}\Sigma^{-1}(x-\mu)}. \tag{10}$$

下面的引理是高斯概率密度函数进行变量的线性变换时的一个标准恒等式。

引理 1. 对于任意可逆的 $d \times d$ 矩阵 A、任意 $\mu, b \in \mathbb{R}^d$，以及任意正定矩阵 $\Sigma \in \mathbb{R}^{d\times d}$，有

$$\phi(x; \mu, \Sigma) = |\det A|\, \phi(Ax + b;\, A\mu + b,\, A\Sigma A^\top). \tag{11}$$

接下来，我们给出在对 $X \sim \mathcal{N}(\mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})$ $Y \sim \mathcal{N}(\mu_{\mathrm{y}}, \Sigma_{\mathrm{y}})$施加相同仿射变换时，$\rho_{X,Y}$ 的一个恒等式，其中 $\rho_{X,Y}$表示对任意 $t \in [0, 1]$，距离 (3) 中的 $S = t\Sigma_{\mathrm{x}} + (1-t)\Sigma_{\mathrm{y}}$。

引理 2. 对于任意可逆的 $d \times d$ 矩阵 A 以及任意 $b \in \mathbb{R}^d$,

$$\rho_{AX+b,AY+b}(Ax + b, Ay + b) = \rho_{X,Y}(x, y). \tag{12}$$

证明。 变换后的高斯分布 $AX + b, AY + b$ 对应的矩阵 S （见 (3)）为 $ASA^\top = t\, A\Sigma_{\mathrm{x}}A^\top + (1-t)\, A\Sigma_{\mathrm{y}}A^\top$。由于 $(Ax + b) - (Ay + b) = A(x - y)$，因此有

$$\begin{aligned}
\rho_{AX+b,AY+b}(Ax + b, Ay + b) &= (Ax - Ay)^\top (ASA^\top)^{-1}(Ax - Ay) \\
&= (x - y)^\top A^\top A^{-\top} S^{-1} A^{-1} A\,(x - y) \\
&= (x - y)^\top S^{-1}(x - y) \\
&= \rho_{X,Y}(x, y).
\end{aligned} \tag{13}$$

□

以下推论由引理 2 直接得出，并用于命题 1。

推论 1. 设 $X \sim \mathcal{N}(\mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})$ 和 $Y \sim \mathcal{N}(\mu_{\mathrm{y}}, \Sigma_{\mathrm{y}})$。令 $S^{\mathrm{tri}}_{X,Y}$和 $S^{\mathrm{duo}}_{X,Y}$分别表示 (6)、(7) 中的集合S_{tri} 和 S_{duo}，其中 $\rho = \rho_{X,Y}$。对于任意可逆的 $d \times d$ 矩阵 A 以及任意 $b \in \mathbb{R}^d$，有

$$S^{\mathrm{tri}}_{AX+b,AY+b} = \{(Ax_1 + b, Ax_2 + b, Ay + b) : (x_1, x_2, y) \in S^{\mathrm{tri}}_{X,Y}\}, \tag{14}$$

$$S^{\mathrm{duo}}_{AX+b,AY+b} = \{(Ax_{\mathrm{ref}} + b, Ax_1 + b, Ay_1 + b) : (x_{\mathrm{ref}}, x_1, y_1) \in S^{\mathrm{duo}}_{X,Y}\}. \tag{15}$$

以下结果确立了 $p_{\mathrm{tri}}, p_{\mathrm{duo}}$ 的仿射不变性，是本节的主要分析结果。

命题 1。 对于任意可逆的 $d \times d$ 矩阵 A 和任意 $b \in \mathbb{R}^d$，当输入的高斯分布分别经过仿射变换 $x \mapsto Ax + b$ 后，函数 p_{tri} 和 p_{duo}（见 （1）、（2））保持不变。也就是说，

$$p_{\mathrm{tri}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma_{\mathrm{x}}, \Sigma_{\mathrm{y}}) = p_{\mathrm{tri}}(A\mu_{\mathrm{x}} + b,\, A\mu_{\mathrm{y}} + b,\, A\Sigma_{\mathrm{x}}A^\top,\, A\Sigma_{\mathrm{y}}A^\top), \tag{16}$$

$$p_{\mathrm{duo}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma_{\mathrm{x}}, \Sigma_{\mathrm{y}}) = p_{\mathrm{duo}}(A\mu_{\mathrm{x}} + b,\, A\mu_{\mathrm{y}} + b,\, A\Sigma_{\mathrm{x}}A^\top,\, A\Sigma_{\mathrm{y}}A^\top). \tag{17}$$

Proof. 我们首先证明公式（16）。将引理 1 代入（4），可得

$$p_{\mathrm{tri}}(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}})=\int_{S^{\mathrm{tri}}_{X,Y}}\phi(Ax_1+b;A\mu_{\mathrm{x}}+b,A\Sigma_{\mathrm{x}}A^\top)\,\phi(Ax_2+b;A\mu_{\mathrm{x}}+b,A\Sigma_{\mathrm{x}}A^\top)\\ \phi(Ay+b;A\mu_{\mathrm{y}}+b,A\Sigma_{\mathrm{y}}A^\top)\,|\det A|^3\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}y. \tag{18}$$

应用变量替换 $\tilde{\mathrm{x}}_1=Ax_1+b$, $\tilde{\mathrm{x}}_2=Ax_2+b$, and $\tilde{\mathrm{y}}=Ay+b$，微分变为 $\mathrm{d}\tilde{\mathrm{x}}_i=|\det A|\,\mathrm{d}x_i$, $\mathrm{d}\tilde{\mathrm{y}}=|\det A|\,\mathrm{d}y$，而根据推论 1，积分域变为$S^{\mathrm{tri}}_{AX+b,AY+b}$。也就是说，

$$p_{\mathrm{tri}}(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}})=\int_{S^{\mathrm{tri}}_{AX+b,AY+b}}\phi(\tilde{\mathrm{x}}_1;A\mu_{\mathrm{x}}+b,A\Sigma_{\mathrm{x}}A^\top)\,\phi(\tilde{\mathrm{x}}_2;A\mu_{\mathrm{x}}+b,A\Sigma_{\mathrm{x}}A^\top)\\ \phi(\tilde{\mathrm{y}};A\mu_{\mathrm{y}}+b,A\Sigma_{\mathrm{y}}A^\top)\,\mathrm{d}\tilde{\mathrm{x}}_1\,\mathrm{d}\tilde{\mathrm{x}}_2\,\mathrm{d}\tilde{\mathrm{y}}. \tag{19}$$

将 (4) 代入 (19)，即可得到 (16)。

通过相同的计算，将 (4) 的被积函数替换为 (5) 的被积函数，即可得到恒等式（17）。利用推论 1，将 (18) 的积分域 $S^{\mathrm{tri}}_{X,Y}$ 替换为$S^{\mathrm{duo}}_{X,Y}$，其在 $x\mapsto Ax+b$ 下的像为$S^{\mathrm{duo}}_{AX+b,AY+b}$。

□

以下技术引理用于命题 3 的证明，以说明双三元概率 (5) 可以化简为一个单变量函数。

引理 3。 设 x_0、$x_1\in\mathbb{R}^d$ 是相互独立的高斯随机变量 $\mathcal{N}(0,\ I)$，并设 $y\sim\mathcal{N}(\mu,I)$。定义$\mathbf{u}=[u_0\ u_1]$ 为

$$u_0=x_1-x_0,\qquad u_1=y-x_0. \tag{20}$$

Then

$$\mathbf{u}\sim\mathcal{N}\left(\begin{bmatrix}0\\ \mu\end{bmatrix},C\right),\qquad C=\begin{bmatrix}2I & I\\ I & 2I\end{bmatrix}. \tag{21}$$

令 $\mathbf{z}=[z_0\ z_1]$，其中 z_0、$z_1\in\mathbb{R}^d$ 由$\mathbf{u}=C^{1/2}\mathbf{z}$ 定义。那么$\mathbf{z}=[z_0\ z_1]\sim\mathcal{N}(m,I)$，其中

$$m=C^{-1/2}[0\ \mu]^\top=\left[\tfrac{\sqrt{3}-3}{6}\mu\ \ \tfrac{3+\sqrt{3}}{6}\mu\right]^\top. \tag{22}$$

证明。由于$\mathbf{u}$ 是高斯分布 x_0,x_1,y 的线性组合，因此$\mathbf{u}$ 也服从高斯分布。根据期望的线性性，$\mathbb{E}[u_0]=0$ 和 $\mathbb{E}[u_1]=\mu$。由于 x_0,x_1,y 相互独立，

$$\mathrm{var}(u_0)=\mathrm{var}(u_1)=2I,\qquad \mathrm{cov}(u_0,u_1)=\mathrm{var}(x_0)=I. \tag{23}$$

直接计算可得

$$C^{1/2}=\begin{bmatrix}\alpha I & \beta I\\ \beta I & \alpha I\end{bmatrix},\qquad \alpha=\frac{\sqrt{3}+1}{2},\quad \beta=\frac{\sqrt{3}-1}{2} \tag{24}$$

因此

$$C^{-1/2} = \frac{1}{\alpha^2 - \beta^2} \begin{bmatrix} \alpha I & -\beta I \\ -\beta I & \alpha I \end{bmatrix} = \begin{bmatrix} \frac{3+\sqrt{3}}{6} I & \frac{\sqrt{3}-3}{6} I \\ \frac{\sqrt{3}-3}{6} I & \frac{3+\sqrt{3}}{6} I \end{bmatrix}. \tag{25}$$

方程（22）可直接由方程（25）以及以下事实得出：

$$\mathbb{E}[z] = \mathbb{E}[C^{-1/2}u] = C^{-1/2}[0\ \mu]^\top. \tag{26}$$

□

3 三角形概率

本节将介绍一种高效算法，用于近似函数 p_{tri} （见（1））。该函数可分解为 p_{tri}，其中 $p_{\text{tri}} = h \circ g$，其中 $g\colon\ \mathbb{R}^{2d+d(d+1)} \to \mathbb{R}^{2d}$ 的计算开销较低，而 $h\colon\ \mathbb{R}^{2d} \to [0,1]$ 则通过一个小型神经网络近似。本节中，ρ 为带有 $S = \frac{2}{3}\Sigma_{\text{x}} + \frac{1}{3}\Sigma_{\text{y}}$ 的（3）。

3.1 将 p_{tri} 化简为 $2d$ 个参数

将 p_{tri} 的维数降低所需的主要数学工具是命题 1，而变换分两步进行。首先，对瑟斯顿模型中的两个高斯分布进行平移，使其中一个的均值位于原点。然后，同时对两个高斯分布的协方差矩阵进行对角化。下面将说明如何将 p_{tri} 从 $2d + d\ (d+1)$ 个参数化简为 $2d$ 个参数。

首先，通过平移两个高斯分布，或给 μ_{x} 和 μ_{y}加上同一个向量，$p_{\text{tri}}(\mu_{\text{x}},\ \mu_{\text{y}},\ \Sigma_{\text{x}},\ \Sigma_{\text{y}})$保持不变。特别地，根据命题 1，我们有

$$p_{\text{tri}}(\mu_{\text{x}}, \mu_{\text{y}}, \Sigma_{\text{x}}, \Sigma_{\text{y}}) = p_{\text{tri}}(0, \mu_{\text{y}} - \mu_{\text{x}}, \Sigma_{\text{x}}, \Sigma_{\text{y}}) \tag{27}$$

对所有 $\mu_{\text{x}},\ \mu_{\text{y}} \in \mathbb{R}^d$ 以及所有 $\Sigma_{\text{x}},\ \Sigma_{\text{y}}$均成立。

接下来，我们同时对角化 (3) 中的 S 和 Σ_{x}。由于 S 是 Σ_{x} 和 Σ_{y}的线性组合，同时对角化 S、Σ_{x}、Σ_{y} 中的任意两个，就能使第三个也对角化。具体而言，我们首先通过变换将 S（见 (3)）化为单位矩阵。这一变换将 (3) 中的距离函数 ρ变为欧几里得距离函数。也就是说，我们用矩阵 $M = L^{-1}$乘以 (27) 右侧的高斯分布[2]，其中 $S = LL^\top$ 是 S的 Cholesky分解，因此 $MSM^\top = I$。注意，任何满足 $MSM^\top = I$的 M都足够。根据命题 1，我们知道这一变换保持 p_{tri}不变。也就是说，

$$p_{\text{tri}}(\mu_{\text{x}}, \mu_{\text{y}}, \Sigma_{\text{x}}, \Sigma_{\text{y}}) = p_{\text{tri}}(0, M(\mu_{\text{y}} - \mu_{\text{x}}), M\Sigma_{\text{x}}M^\top, M\Sigma_{\text{y}}M^\top). \tag{28}$$

接下来，我们对 p_{tri} 的其余高斯分布自变量进行变换，使其协方差矩阵变为对角矩阵。利用特征分解

$$M\Sigma_{\text{x}}M^\top = V\Lambda V^\top, \tag{29}$$

[2]在本节中，为了便于说明，我们将 p_{tri} 称为两个高斯分布的函数。也就是说，$p_{\text{tri}}(\mu_{\text{x}}, \mu_{\text{y}}, \Sigma_{\text{x}}, \Sigma_{\text{y}})$ 是高斯分布 $\mathcal{N}(\mu_{\text{x}}, \Sigma_{\text{x}})$ 和 $\mathcal{N}(\mu_{\text{y}}, \Sigma_{\text{y}})$ 的函数。

我们通过下式对（28）右侧的高斯分布进行变换:

$$p_{\mathrm{tri}}(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}})=p_{\mathrm{tri}}(0,V^{\top}M(\mu_{\mathrm{y}}-\mu_{\mathrm{x}}),V^{\top}M\Sigma_{\mathrm{x}}M^{\top}V,V^{\top}M\Sigma_{\mathrm{y}}M^{\top}V). \tag{30}$$

式（30）的协方差通过以下恒等式相关联

$$\tfrac{2}{3}\Sigma_{\tilde{\mathrm{x}}}+\tfrac{1}{3}\Sigma_{\tilde{\mathrm{y}}}=I,\qquad \Sigma_{\tilde{\mathrm{x}}}=V^{\top}M\Sigma_{\mathrm{x}}M^{\top}V,\qquad \Sigma_{\tilde{\mathrm{y}}}=V^{\top}M\Sigma_{\mathrm{y}}M^{\top}V \tag{31}$$

并且 $\Sigma_{\tilde{\mathrm{x}}}$，$\Sigma_{\tilde{\mathrm{y}}}$ 二者都是对角矩阵。综上，将仿射变换

$$x\mapsto V^{\top}M(x-\mu_{\mathrm{x}}), \tag{32}$$

应用于 $p_{\mathrm{tri}}(\mu_{\mathrm{x}}$，$\mu_{\mathrm{y}}$，$\Sigma_{\mathrm{x}}$，$\Sigma_{\mathrm{y}})$ 的高斯分布，并利用恒等式（31），可得

$$p_{\mathrm{tri}}(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}})=p_{\mathrm{tri}}(0,V^{\top}M(\mu_{\mathrm{y}}-\mu_{\mathrm{x}}),\Sigma_{\tilde{\mathrm{x}}},3I-2\Sigma_{\tilde{\mathrm{x}}}) \tag{33}$$

其中 $\Sigma_{\tilde{\mathrm{x}}}$ 是对角协方差矩阵。

注 1。 在不存在重特征值的情况下，特征分解 （29）仅由特征向量（ V 的列）的符号决定；重特征值只会在一个零测集上出现。将 V 的第 j 列乘以 -1 会使 $\Sigma_{\tilde{\mathrm{x}}}$、$\Sigma_{\tilde{\mathrm{y}}}$ 保持不变，但会改变式（*33*）的自变量 $V^{\top}M$（$\mu_{\mathrm{y}}-\mu_{\mathrm{x}}$）的第 j 个分量的符号。因此，我们任意选择V，使均值的每个分量均为非负。等价地，我们可以对向量 $V^{\top}M$（$\mu_{\mathrm{y}}-\mu_{\mathrm{x}}$）的每个分量取绝对值。

注释 2. 我们对 Λ 的对角元素 $\lambda_1,\ldots,\lambda_d$ 进行排序，并相应地调整 V 的列，使得 $\lambda_1\geq\cdots\geq\lambda_d$。对于任意置换矩阵 P，矩阵 $P\Lambda P^{\top}$ 是由置换后的特征值构成的对角矩阵，因此根据命题 *1*，对 Λ 的元素进行置换并相应地调整 V 的列，不会改变 p_{tri} 。

经过上述变换后， p_{tri} 成为 $2d$ 个参数 (λ,η) 的函数，其中 $\lambda\in\mathbb{R}^d$ 是 Λ 的对角线，并且我们将 η 定义为 $\eta=|V^{\top}M(\mu_{\mathrm{y}}-\mu_{\mathrm{x}})|\in\mathbb{R}^d_{\geq 0}$，其中绝对值是逐元素取的。为方便记号，我们用 $g:\mathbb{R}^{2d+d(d+1)}\to\mathbb{R}^{2d}$ 表示将 $(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}})$ 映射到 (λ,η) 的函数，并将 h 定义为函数

$$h(\lambda,\eta)=h(g(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}}))=p_{\mathrm{tri}}(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}}). \tag{34}$$

下面我们证明，对p_{tri}的自变量进行 g 变换后，函数 h 的自变量为 $\lambda\in(0,\ 3\,2)^d$和 $\eta_i\geq 0$，其中 $i=1,\ .....,\ d$。

断言2。令

$$D_d=\{(\mu_{\mathrm{x}},\mu_{\mathrm{y}},\Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}}):\mu_{\mathrm{x}},\mu_{\mathrm{y}}\in\mathbb{R}^d,\ \textit{and symmetric positive definite}\ \Sigma_{\mathrm{x}},\Sigma_{\mathrm{y}}\in\mathbb{R}^{d\times d}\}$$

and let

$$\Omega_d=\left\{\lambda\in\mathbb{R}^d:\tfrac{3}{2}>\lambda_1\geq\cdots\geq\lambda_d>0\right\}\times\mathbb{R}^d_{\geq 0}. \tag{35}$$

那么，*(34)*的 g 将 D_d 映射到Ω_d。

证明。 我们首先证明 $g(D_d)\subseteq\Omega_d$。(34)中的向量 λ 是 $\Sigma_{\tilde{\mathrm{x}}}$的对角线，并且根据(31)， $\Sigma_{\tilde{\mathrm{x}}}=(3I-\Sigma_{\tilde{\mathrm{y}}})/2$ 。 $\Sigma_{\tilde{\mathrm{x}}}$ 的正定性蕴含 $\lambda_i>0$ ，而 $\Sigma_{\tilde{\mathrm{y}}}$的正定性给出：对于 $i=1,\ .....,\ d$，有 $\lambda_i<3/2$。由于根据注释1和2， $\eta_i\geq 0$且 $\lambda_1\geq\cdots\geq\lambda_d$，因此我们有 $g(D_d)\subseteq\Omega_d$。

现在我们证明 $\Omega_d\subseteq g(D_d)$。令 $(\lambda,\ \eta)\in\Omega_d$，并设 $\mu_{\mathrm{x}}=0$、$\mu_{\mathrm{y}}=\eta$、$\Sigma_{\mathrm{x}}=\Lambda$ 和 $\Sigma_{\mathrm{y}}=3I-2\Lambda$。那么 $g(\mu_{\mathrm{x}},\ \mu_{\mathrm{y}},\ \Sigma_{\mathrm{x}},\ \Sigma_{\mathrm{y}})=(\lambda,\ \eta)$。 □

3.2 h的神经网络近似

经过上一节所述的 g 变换后，现在我们可以通过计算函数 $h:\Omega_d \to [0,1]$ 来评估 p_{tri}。也就是说，我们已将高效评估 p_{tri} 的问题（它是一个具有 $2d + d(d+1)$ 个自变量的函数），转化为评估 h 的问题（它是一个具有 $2d$ 个自变量的函数）。遗憾的是，对于维数大于一的刺激空间，$2d$ 维仍然通常过高，无法使用标准张量积求积规则高效评估 h；这类规则的计算成本按 $\mathcal{O}(n^{2d})$ 增长，其中 n 是单个维度上的节点数。为避免这种指数级计算成本，我们用一个小型神经网络近似 h。事实证明，小型经典神经网络能够以合理的准确度近似 h；训练完成后，与准确度相当的蒙特卡洛方法相比，这些神经网络的评估速度可快几个数量级。

我们使用通过蒙特卡洛计算 p_{tri} 得到的带标签数据训练神经网络。我们通过以下流程在 Ω_d 中选择训练点。首先在 $[10^{-3}, \frac{3}{2} - 10^{-3}]^d$ 上均匀采样 λ_i，然后对这些点排序。[3] 接着，我们先在限制于正正交象限的单位球面上均匀采样方向 $\eta/\|\eta\|$，以此采样 η_i。然后在 $(0, R)$ 上均匀采样范数 $\|\eta\|$，其中 $R > 0$ 按如下方式选取。我们选择 R，使得对于所有 $\lambda_i \in (0, 3/2)$ 以及 $\|\eta\| > R$，$h(\lambda, \eta)$ 都能以 1 为良好近似。

我们利用以下两个观察结果构造一个数值过程来选择 R。首先，根据经验观察，对于固定的 λ，h 是关于每个 η_i 的递增函数。此外，对于固定的 $\|\eta\|$，当 $\gamma > 0$ 较小时，$1 - h(\lambda, \eta)$ 在 η 和 $\lambda \in [\gamma, \frac{3}{2} - \gamma]^d$ 上的最大值于 $\lambda = (\frac{3}{2} - \gamma)(1, \ldots, 1)$ 处取得。接着，我们观察到，对于各分量均为常数的 λ，$h(\lambda, \eta)$ 在对 η 进行正交变换后保持不变。

利用这两个观察结果，我们通过以下机制选择 R。我们设定 $\lambda = (\frac{3}{2} - 10^{-3}, \ldots, \frac{3}{2} - 10^{-3})$ 和 $\eta = \|\eta\|(1, 0, \ldots, 0)$，其中，由于对于具有常数分量的 λ，$h(\lambda, \eta)$ 仅通过 $\|\eta\|$ 依赖于 η，因此 η 的角分量可以任意选取。然后，我们通过蒙特卡洛方法，在 $\|\eta\|$ 上的一维网格中计算 $h(\lambda, \eta)$。对于某个误差容限 δ，我们计算满足 $1 - h(\lambda, \eta) < \delta$ 的最小 $\|\eta\|$。我们发现，对于 $d = 2$，$\|\eta\| = 10.9$ 满足 $\delta = 10^{-4}$；对于 $d = 3$，$\|\eta\| = 11.1$ 满足 $R = 12$，并对两者采用相同的设定。我们证明，对于所有 $\|\eta\| \geq R$ 和 $\lambda_i \in [10^{-3}, \frac{3}{2} - 10^{-3}]$，$d = 2$ 的网络输出与 1 的差值不超过 2.2×10^{-5}。[4]

我们使用 p_{tri} 评估包含 10^5 蒙特卡洛样本的训练数据。我们使用的训练集由 3×10^5 个数据点组成。每个 3×10^5 训练标签的误差标准差为 $\sqrt{h(1-h)/n}$ 其中 h 是真实值，n 是蒙特卡洛样本的数量。因此，最大标准差的上界为 0.002。令 $\|\eta\|$ 在 $(0, R)$ 上服从均匀分布，训练集中的数据点在原点附近的集中程度高于 R 附近。具体而言，单位体积密度与 $\|\eta\|^{1-d}$ 成正比。我们之所以做出这一选择，是因为当 $\|\eta\|$ 接近 R 时，函数 p_{tri} 接近 1。我们通过经验观察发现，采用这一训练分布时，验证集上的误差小于在球体 $\|\eta\| \leq R$ 上均匀采样的 η 的误差。

[3] 当 $\lambda_i \to 0$ 时，Σ_x 的条件数趋于无穷大；当 $\lambda_i \to \frac{3}{2}$ 时，Σ_y 的条件数趋于无穷大。这些情况通常与我们所考虑的应用无关。因此，我们将 $\lambda_i \in [\epsilon, \frac{3}{2} - \epsilon]$ 限制为满足 $\epsilon = 10^{-3}$。[4] 证明使用形式化验证软件 Super [41]，并依赖于两个结果的结合。其中一个给出了 $R \leq \|\eta\| \leq 1000$ 的界。另一个证明网络输出的对数几率至少为 $\|\eta\| - 323$，因此对于 $\|\eta\| > 1000$，该界成立。这两个结果都位于代码仓库 https://github.com/pgree/fast-likelihood-sensory-discrimination 中。

我们用于 $d = 2$,3的网络架构是一个多层感知机，其各层宽度为 $2d$、16、16、1，采用 GELU 激活函数和逻辑输出单元。网络的输入为（log λ, η），并标准化为零均值和单位方差。我们通过经验观察发现，当使用 λ作为输入而不是 log λ时，大误差集中在$\lambda = 0$附近。该网络针对 $d = 2$ 包含 369 个参数，针对 $d = 3$包含 401 个参数。我们通过使用 Adam [30], 、批量大小为 1024 训练网络，经过 1.12×10^5 个步骤，并按照余弦调度将学习率从 10^{-3} 开始衰减；最后根据训练标签上的经验风险，从三次重新启动中保留表现最佳的一次。

我们按如下方式衡量网络的准确度。我们在 4992 个未用于训练的点上对其进行评估，其中 $\lambda_i \in [10^{-3}, \frac{3}{2} - 10^{-3}]$和 $\|\eta\| \leq 12$。测试集由 4992 个点组成，这些点经过挑选，使得在测试集上评估的 h 大致服从 $[\frac{1}{3}, 1]$ 上的均匀分布。我们首先从训练点的分布中抽取 12×4992 个候选点。然后将 $[\frac{1}{3}, 1]$ 划分为 24 个等长区间，并从每个区间中不放回地抽取 208 个样本，共得到 $24 \times 208 = 4992$ 个点。在测试集的每个点上，我们使用 $n = 10^7$ 个样本，通过蒙特卡洛计算 h。这些参考值的标准差至多为 1.6×10^{-4}。我们发现，对于 $d = 2$，网络与参考值之间的均方根差为 7.9×10^{-4}；对于 $d = 3$，均方根差为 9.3×10^{-4}，相应的标准误分别为1.7×10^{-5} 和 1.8×10^{-5}；在这 4992 个点上的最大误差分别为 6.6×10^{-3} 和 7.0×10^{-3}。使用 $n = 10^5$ 个样本计算得到的训练标签，其均方根标准差为 8.3×10^{-4}，因此，网络的误差与训练标签测量时的准确度相当。

3.3 基于原始参数的神经网络

在本节中，到目前为止，我们描述了一种通过函数组合 $h \circ g$ 来近似 p_{tri} 的方法，其中 g 执行降维，而 h 则通过神经网络进行近似。该方法的一个自然替代方案是，使用一个网络来近似 p_{tri}，其输入为公式 (1) 中的 $d + d(d + 1)$ 参数（利用均值的平移不变性，从 p_{tri} 中移除 d 参数）。

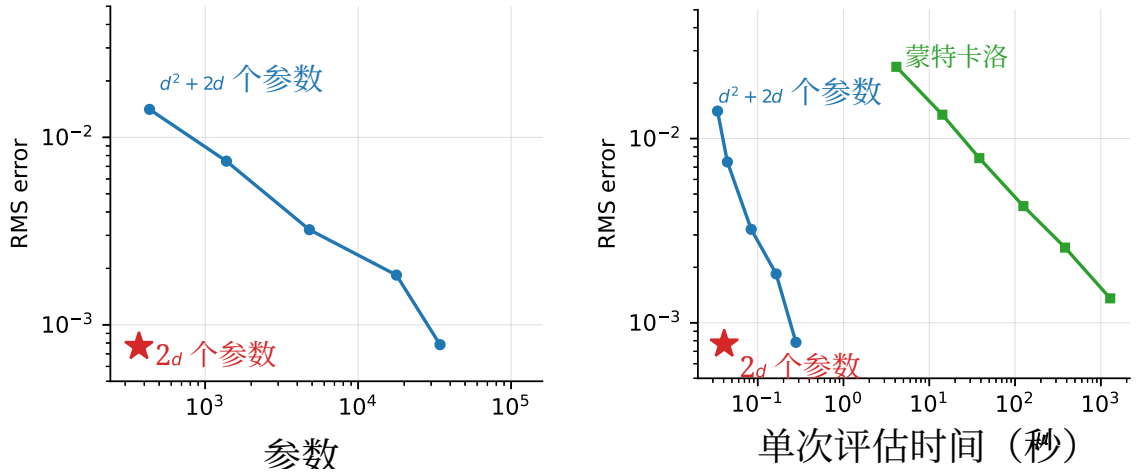
我们使用 p_{tri} 的原始 $d + d(d + 1)$ 参数训练了不同规模的神经网络，并将其性能与以下两种方法进行比较：(i) 在降维后的 h 上训练的神经网络；(ii) 蒙特卡洛。在图3中，我们根据网络参数数量，将两个神经网络的准确度作为函数进行比较（图3a）；在图3b中，我们比较了两个网络和蒙特卡洛方法在单次评估中准确度随时间的变化。图3使用 $d = 2$，误差是对验证集上4992个点计算的均方根差；这些点以与测试集相同的方式进行采样和标注。

对于较大的网络，我们训练了深度为 2、宽度分别为 16、32、64 和 128 的网络，以及深度为 3、宽度为 128 的网络。较大的网络需要多两个数量级的参数才能达到相同的准确度，评估所需时间也约为七倍。需要指出的是，所有参数数量相同的网络都使用相同的训练数据进行训练。我们用于大型网络（即使用原始$2d + d^2$ 参数进行训练的网络）的训练数据，是通过对小型网络（即为逼近 h 而训练的网络）所用的训练数据进行变换得到的。具体而言，对于 $(\lambda, \eta) \in \Omega_d$，根据断言 2，原像 $g^{-1}(\lambda, \eta) = \{(\mu_x, \mu_y, \Sigma_x, \Sigma_y) \in D_d : g(\mu_x, \mu_y, \Sigma_x, \Sigma_y) = (\lambda, \eta)\}$ 非空，我们按如下方式随机选取其中一个元素。令 $A = QD_s$，其中 Q 是一个旋转

从 $SO(d)$ 中均匀抽取，且对角矩阵D_s每个对角元素的 $\log_{10}$ 相互独立，并在区间 $(-\frac{1}{4},\ _{14})$ 上均匀分布。接着令 $\mu_x = 0, \mu_y = A\eta, \Sigma_{\bar{x}} = A\,\mathrm{diag}(\lambda)A^\top$，并令 $\Sigma_y = A(3I - 2\mathrm{diag}(\lambda))A^\top$。

备注 3. 针对原始参数训练的神经网络，其训练集的构建依赖于变量约简。训练点取自 $2d$ 维集合 Ω_d，而非原始参数所在的 $(d + d(d+1))$ 维空间。若每个维度上取 n 个点，张量积网格将有 $n^{d+d(d+1)}$ 个点，而不是 n^{2d} 个。

计时结果为 Apple M3 Max 笔记本电脑中央处理器上的墙钟时间，使用 JAX 0.10.0 [4], ，不包括编译时间。神经网络在每批 20，000 次试次上计时。对于 $d = 2$，只要特征分解 (29) 以闭式计算，将 p_{tri} 维数约简的函数 g 与神经网络合计每次评估耗时 $0.041\,\mu$ 秒。在准确度相同的情况下，使用原始参数的神经网络耗时 $0.28\,\mu$ 秒。



(a) 参数数量　　(b) 单次评估时间

图3： p_{tri} 单次评估的准确度随参数数量（图3a）和评估时间（图3b）的变化。我们比较了在具有原始参数数量的 p_{tri} 上训练的不同规模神经网络（蓝色），以及在 h 的 $2d$ 输入上训练的神经网络（红色）（见（34））。在图3b（评估时间）中，我们将这两种神经网络方法与 [22] 的蒙特卡洛方法（绿色）进行比较。对于 $d = 2$，误差为均方根差；评估基于一组包含4992个点的数据集，其参考值通过使用 $n = 10^7$ 个样本的蒙特卡洛计算得到。

3.4 快速统计推断算法

颜色辨别模型 [22] 采用一种实验程序，在二维刺激空间——即所谓的等亮平面上进行三角测试（示例见图 1）。我们将贝叶斯模型拟合到实验试验数据。实验包含来自 [22] 的参与者 CH 的 $N = 6000$ 次试次。在每次试次中 $i = 1, \ldots\ldots, N$,

受试者会看到重复刺激 $\mu_{\mathrm{x},\,i}$ 两次，以及独特刺激 $\mu_{\mathrm{y},\,i}$ 一次；$y_i = 1$ 表示回答正确，$y_i = 0$ 则表示回答错误。该模型

$$
\begin{aligned}
y_i \mid \theta &\sim \mathrm{Bernoulli}\big(p_i(\theta)\big), && (36)\\
p_i(\theta) &= p_{\mathrm{tri}}\big(\mu_{\mathrm{x},i}, \mu_{\mathrm{y},i}, \Sigma(\mu_{\mathrm{x},i}, \theta), \Sigma(\mu_{\mathrm{y},i}, \theta)\big), && (37)\\
\Sigma(x, \theta) &= U(x, \theta)\, U(x, \theta)^{\top}, && (38)\\
U_{ab}(x, \theta) &= \sum_{k,\ell=0}^{4} \theta_{k\ell ab}\, T_k(\xi_1)\, T_\ell(\xi_2),\\
\theta_{k\ell ab} &\sim \mathcal{N}\big(0,\, 3 \times 10^{-4} \cdot 0.4^{k+\ell}\big).
\end{aligned}
$$

这里，T_k 是次数为 k的切比雪夫多项式，$x = (\xi_1,\ \xi_2) \in [-1, 1]^2$、$a \in \{1, 2\}$ 和$b \in \{1, 2, 3\}$，因此 $U(x,\ \theta)$是一个 2×3 矩阵。在给定 θ的条件下，y_i 相互独立，且 $\theta_{k\ell ab}$相互独立。在推断中，[22] 使用最大后验估计。也就是说，他们求解

$$
\hat{\theta} = \arg\max_{\theta} \Big(\sum_{i=1}^{N} \big[y_i \log p_i(\theta) + (1 - y_i) \log\big(1 - p_i(\theta)\big) \big] + \log p(\theta) \Big), \tag{39}
$$

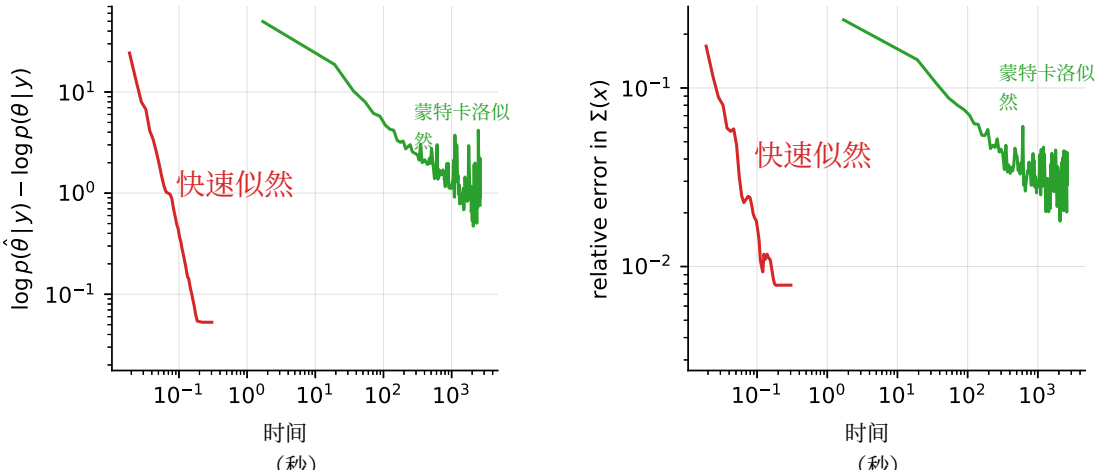
其中，$p(\theta)$是先验密度。他们使用蒙特卡罗估计来计算三角形概率p_{tri}。

这里，我们使用 p_{tri} 的快速似然近似，并同样采用最大后验估计来拟合相同的模型。我们在中央处理器上比较这两种方法的耗时和准确度。图4展示了两组耗时随准确度变化的对比。图4a展示了$\hat{}\log\ p(\theta|y) - \log\ p(\theta|y)$随时间的变化，其中 $p(\theta|y)$是通过快速似然计算得到 p_{tri}的后验分布。图4b展示了相对误差

$$
\left(\frac{1}{m} \sum_{j=1}^{m} \frac{\|\Sigma(x_j, \theta) - \Sigma(x_j, \hat{\theta})\|_F^2}{\|\Sigma(x_j, \hat{\theta})\|_F^2} \right)^{1/2} \tag{40}
$$

在笛卡尔网格上对一组点 x_1、……、x_m 处的协方差场进行近似时，θ 是各方法当前的迭代点，θ 则是已知的最佳后验最大后验估计，使用快速似然计算得出，是从独立先验抽样中进行的 200 次 L-BFGS-B 拟合中最优的结果。两种拟合都从 $\theta + 0.2\,\xi$开始，其中 ξ 是一次先验抽样。图 4 中的曲线从各方法的第一次迭代后开始。使用快速似然时，L-BFGS-B 在 0.30 秒后得到的结果为 $\hat{}\log\ p(\theta|y)\ - \log\ p(\theta|y)\ = 0.053$，不包括 0.21 秒的编译时间，在 Σ中的相对误差为 0.008。按照 [22], 的方法，我们还使用随机梯度下降和动量，通过蒙特卡洛似然近似进行 1500 步拟合[5] 。拟合在 2618 秒后得到的结果为 $\hat{}\log\ p(\theta|y)\ - \log\ p(\theta|y)\ = 0.76$，相对误差为 0.029。若在对数后验中改用蒙特卡洛方法评估 p_{tri}，并对每个试次使用 10^5 个样本，则两种拟合的最终结果 $\hat{}$ 分别比 θ 低 0.05 和 0.85，标准误分别为 0.07 和 0.08。

[5]在此次似然评估中，用 $1/(1 + e^{t/b})$ 中公式 (4) 的 (u_1，u_2，v) 的 2000 个样本的均值近似 p_{tri}，其中 $t = \rho(u_1,\ u_2) - \min\{\rho(u_1,\ v)\ \rho(u_2,\ v)\}$ 和 $b = 5 \times 10^{-3}$ 替代了 S_{tri} 的指示函数，从而使 p_{tri} 关于 θ 可微。



(a) 目标函数差异　　(b) 协方差场差异

图 4：模型在 [22],处的最大后验估计所对应的时间与准确度比较，其中 p_{tri} 分别通过蒙特卡洛似然（绿色）和快速似然 ˆ（红色）计算。我们展示了 (a) 最大后验处的对数后验 θ减去当前θ处的对数后验，以及 (b) 当前参数 θ 下 $\Sigma(x)$ 相对于最大后验估计处 $\Sigma(x)$ 的相对误差。最大后验 θ 是使用快速似然进行 200 次 L-BFGS-B 拟合后所得的最佳结果。使用 $n = 10^5$ 进行蒙特卡洛计算时，(a) 中最终的差值相同，误差在 0.09 以内。

4 二选一概率

本节介绍一种高效计算二-三点检验伯努利概率p_{duo} 的算法（见式（5））。我们假设，对于任意两个呈现的刺激 $\Sigma_{\text{x}} = \Sigma_{\text{y}}$ x、 $y \in \mathcal{X}$，[6]。对于欧几里得 ρ，当 $\Sigma_{\text{x}} \neq \Sigma_{\text{y}}$ 时的情况已在 [10], 中处理，其形式与式（3）一致，只需将 S 变换为单位矩阵。下文中，当 $\Sigma_{\text{x}} = \Sigma_{\text{y}}$时，二选一概率可化为一个变量的解析函数。对 [10] 进行特征分解和级数求和，将由切比雪夫展开的计算取代；其系数在预计算期间计算一次，精度达到用户指定的精度 ϵ。

在我们针对二-三点检验所考虑的瑟斯顿模型中，有三个高斯分布：x_0， x_1 独立同分布于 $\mathcal{N}(\mu_{\text{x}},\ \Sigma)$，并且与 $y \sim \mathcal{N}(\mu_{\text{y}},\ \Sigma)$ 相互独立。以下是我们算法背后的核心数学观察。我们证明， p_{duo}完全通过标量 $r = (\mu_{\text{y}} - \mu_{\text{x}})^{\top}\Sigma^{-1}(\mu_{\text{y}} - \mu_{\text{x}})$ 依赖于 μ_{x}， μ_{y}， Σ。此外，p_{duo}作为 r的函数是解析的，并且可以通过计算上可行的一维积分求得。

命题 3。 对于所有 μ_{x}、 $\mu_{\text{y}} \in \mathbb{R}^d$，以及所有正定对称的 $d \times d$ 矩阵 Σ,

$$p_{\text{duo}}(\mu_{\text{x}}, \mu_{\text{y}}, \Sigma, \Sigma) = \Pr\big(\chi'^2_d(\lambda_1) > \chi'^2_d(\lambda_0)\big) \tag{41}$$

[6]这一假设并不意味着拟合协方差场是常数。相反，我们可以将这一假设视为一种近似；当 x趋近于 y时，其误差趋近于零。

其中

$$\lambda_1 = \tfrac{r}{6}\big(2+\sqrt{3}\big), \quad \lambda_0 = \tfrac{r}{6}\big(2-\sqrt{3}\big), \quad r = (\mu_{\mathrm{y}}-\mu_{\mathrm{x}})^{\top}\Sigma^{-1}(\mu_{\mathrm{y}}-\mu_{\mathrm{x}}) \tag{42}$$

其中， $\chi'^2_d(\lambda)$ 表示自由度为 d 、非中心参数为 λ 的非中心卡方分布。

证明。我们首先将命题 1 应用于 $p_{\mathrm{duo}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma, \Sigma)$，其中 $A = \Sigma^{-1/2}$ 和 $b = -\Sigma^{-1/2}\mu_{\mathrm{x}}$，得到

$$p_{\mathrm{duo}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma, \Sigma) = p_{\mathrm{duo}}(0, \mu, I, I) \tag{43}$$

其中

$$\mu = \Sigma^{-1/2}(\mu_{\mathrm{y}} - \mu_{\mathrm{x}}). \tag{44}$$

因此，不失一般性，我们可以假设 $\mu_{\mathrm{x}} = 0$、 $\mu_{\mathrm{y}} = \mu$、 $\Sigma = I$。我们定义u_0, u_1 为以下向量

$$u_0 = x_1 - x_0, \qquad u_1 = y - x_0 \tag{45}$$

并将多元高斯分布 $\mathbf{u} \in \mathbb{R}^{2d}$ 定义为 $\mathbf{u} = [u_0\ u_1]$。于是

$$\mathbf{u} \sim \mathcal{N}\left(\begin{bmatrix}0\\ \mu\end{bmatrix}, C\right), \qquad C = \begin{bmatrix}2I & I\\ I & 2I\end{bmatrix} \tag{46}$$

其中 I 是 $d \times d$ 单位矩阵。令 Q 为标量

$$Q = \|u_0\|^2 - \|u_1\|^2 \tag{47}$$

并注意到

$$p_{\mathrm{duo}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma, \Sigma) = \Pr(Q < 0). \tag{48}$$

将 Q写成二次型

$$Q = \mathbf{u}^{\top} M \mathbf{u}, \qquad M = \begin{bmatrix}I & \\ & -I\end{bmatrix} \tag{49}$$

并利用引理 3，我们知道

$$Q = \mathbf{z}^{\top} C^{1/2} M C^{1/2} \mathbf{z} \tag{50}$$

其中 $\mathbf{z} \sim \mathcal{N}(m, I)$， m 定义见式（22）。结合式（49）和（24），我们注意到二次型（50）的矩阵是对角矩阵。具体而言，

$$C^{1/2} M C^{1/2} = \sqrt{3}\begin{bmatrix}I & \\ & -I\end{bmatrix}. \tag{51}$$

定义$\mathbf{z} = [z_0\ z_1]$ 并将（51）代入（50），我们得到

$$Q = \mathbf{z}^{\top} C^{1/2} M C^{1/2} \mathbf{z} = \sqrt{3}\big(\|z_0\|^2 - \|z_1\|^2\big). \tag{52}$$

由于 z_0，z_1 是协方差为单位矩阵的相互独立的高斯随机变量，$\|z_0\|^2$，$\|z_1\|^2$ 是相互独立的非中心卡方分布，其 d 自由度对应的非中心参数为（见（22））,

$$\|\mathbb{E}[z_1]\|^2 = \|\mu\|^2\Big(\tfrac{3+\sqrt{3}}{6}\Big)^2 = \|\mu\|^2\frac{2+\sqrt{3}}{6}, \qquad \|\mathbb{E}[z_0]\|^2 = \|\mu\|^2\Big(\tfrac{\sqrt{3}-3}{6}\Big)^2 = \|\mu\|^2\frac{2-\sqrt{3}}{6}. \tag{53}$$

将 (48)、(52)、(53) 结合起来，再利用 $\|\mu\|^2 = r$，即可得出方程 (41)。 □

由于方程 (41) 中的卡方分布相互独立，联合分布 $\left(\chi'^2_d(\lambda_1)\ \chi'^2_d(\lambda_0)\right)$ 的概率密度函数是两个一维概率密度函数的乘积。因此,

$$p_{\mathrm{duo}}(r) = \int_0^\infty \int_t^\infty f_{d,\lambda_1(r)}(t')\, f_{d,\lambda_0(r)}(t)\, dt'\, dt = \int_0^\infty \overline{F}_{d,\lambda_1(r)}(t)\, f_{d,\lambda_0(r)}(t)\, dt \tag{54}$$

其中$f_{d,\lambda}$ 表示 $\chi'^2_d(\lambda)$ 的概率密度函数（例如，参见 [27]），其定义为

$$f_{d,\lambda}(t) = \frac{1}{2}e^{-(t+\lambda)/2}\Big(\frac{t}{\lambda}\Big)^{\frac{d}{4}-\frac{1}{2}} I_{\frac{d}{2}-1}\big(\sqrt{\lambda t}\big), \tag{55}$$

其中，I_ν 是第一类修正贝塞尔函数 [7, (10.25.2)], ，而 $F_{d,\lambda}$ 定义为

$$\overline{F}_{d,\lambda}(t) = \int_t^\infty f_{d,\lambda}(\tau)\, \mathrm{d}\tau, \tag{56}$$

这就是所谓的 Marcum Q-函数 $Q_{d/2}(\sqrt{\lambda}, \sqrt{t})$ [33]。高效计算 $f_{d,\lambda}$ 和 $F_{d,\lambda}$ 的算法已在标准软件包中实现（例如，参见 [18, 3]）。因此，对于任意参数，高效计算 $p_{\mathrm{duo}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma, \Sigma)$ 的问题已转化为计算一维积分 (54)。使用自适应求积规则可以将积分 (54) 计算至浮点精度（例如，参见 [44]）。

4.1 数值计算

此处我们介绍一种数值过程，用于在误差容限为 $\epsilon > 0$ 的情况下高效计算 p_{duo}。我们首先注意到，当 $r \to \infty$ 时，$p_{\mathrm{duo}}(r) \to 1$，并选择截断 R，使得对于所有 $r > R$，$|p_{\mathrm{duo}}(r) - 1| < \epsilon/10$。然后，我们在区间 $[0, R]$ 上构造阶数为 n 的切比雪夫展开，使得到第 n 个系数时，系数已经衰减到小于 $\epsilon/10$ 的幅度。由于 p_{duo} 是解析的，切比雪夫展开中的系数呈指数衰减 [44]，因此，在固定 R 的情况下，要达到 ϵ 的准确度，需要 $\mathcal{O}(-\log \epsilon)$ 项。具体而言，我们选择切比雪夫展开的阶数 n，然后对 $i = 1$、.....、n 计算 $p_{\mathrm{duo}}(x_i)$，其中 x_1、.....、x_n 是经过适当变换到区间 $[0, R]$ 上的 n 阶切比雪夫节点 [44]。我们使用自适应 Gauss–Kronrod 求积法 [39, 45], ，以容限 $\epsilon/10$ 计算 $p_{\mathrm{duo}}(x_i)$。

然后，我们计算切比雪夫展开中相应的系数。如果这些系数的绝对值尚未衰减至低于 $\epsilon/10$，我们就使用更大的 n 并重复

这一过程。我们存储切比雪夫展开的 n系数。随后，我们可以通过在任意 $r \in [0, R]$处计算同一个 n阶切比雪夫展开，对任意输入求得 p_{duo}，并达到 ϵ准确度。这一过程能够快速且准确地计算p_{duo}。这种计算适用于自动微分和即时编译[4]。

在图 5 中，我们比较了使用我们的数值过程对 p_{duo} 、 $d = 3$ 和 $\epsilon = 10^{-8}$ 进行评估的结果与采用不同样本数量的蒙特卡洛方法的结果。我们报告了 $[0, R]$ 中 400 个点上的均方根误差，其中真值是使用自适应求积以机器精度计算得到的。切比雪夫展开每次评估耗时 $0.028\,\mu$ 秒，在 $[0, R]$ 上的无穷范数误差为 8.7×10^{-12}，该误差采用自收敛意义进行测量。也就是说，我们使用项数较多且系数计算达到接近浮点精度的切比雪夫展开作为参考。在容差为 10^{-10} 时，使用 15 点高斯–克朗罗求积的自适应求积每次评估耗时 28 毫秒 [39]。 [10,式 (11)] 的方法同样避免了对 p_{duo} 的每次评估都进行求积。然而，达到误差 ϵ 所需的项数会随 r 增长，因为非中心参数 (53) 与 r 成正比。在 $r = R$ [10] 时，要使误差至多为 10^{-8}，该方法需要 2328 项，而本节算法只需 36 项。在 $\epsilon = 10^{-8}$ 时，我们的方法分别给出 $R = 220.8$ 和 $n = 36$ 项，以达到 $d = 3$。对于 $d = 2$，我们得到 $R = 218.4$ 和 $n = 36$。图 5 中的计时不包括预计算步骤；对于给定的 d 和 ϵ，该步骤只需执行一次。

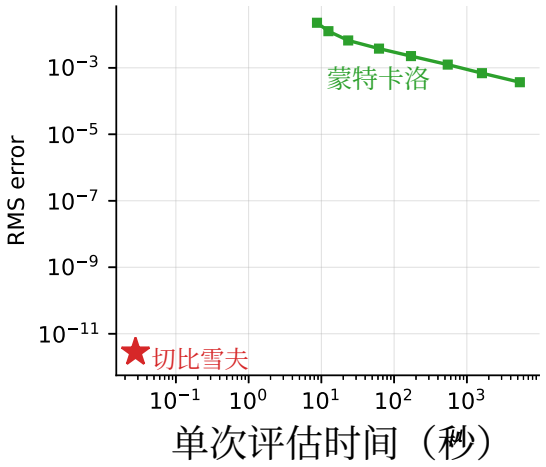


图 5：评估 p_{duo} 对于 $d = 3$的均方根误差。我们将切比雪夫展开与 $\epsilon = 10^{-8}$ （红色）以及蒙特卡洛近似进行比较，针对范围从 30 到 10^5 （绿色）的 n 。

5 讨论

我们介绍了用于计算（1）、（2）的伯努利概率 p_{tri} 和 p_{duo} 的算法，其效率远高于现有方法。我们的方法能够高效地进行统计推断，并完成与瑟斯顿感知模型相关的其他计算任务。例如，在 [22],的贝叶斯模型上使用我们对 p_{tri} 的近似时，与通过蒙特卡洛计算相比，我们能以短三个数量级以上的时间获得更准确的最大后验估计

。由于这些近似方法高效、确定且可微，因此也使得使用基于梯度的采样器（如无转向采样器（NUTS） [21], ）对后验分布进行采样变得切实可行，并能够自适应地安排试次；这需要在试次间隔期间进行后验推断。

对于 p_{tri}， p_{duo}，高效评估所依据的核心数学观察有两点：(i) 将关于 $\mathcal{O}(d^2)$ 个参数的函数解析变换为关于 $\mathcal{O}(d)$ 个参数的函数；(ii) 低维函数是光滑的，可以使用神经网络（p_{tri}）或特殊函数展开（p_{duo}）进行高效近似。对于 p_{tri}，性能提升同时来自变换和网络——在原始参数上，神经网络需要多出两个数量级的参数，并且要多花大约七倍的时间才能达到相同的准确度（图3）。

我们所采用的变量约简依赖于模型的两个性质：(i) 噪声族（此处为高斯分布）在刺激空间的仿射变换下保持封闭；(ii) 响应仅通过形如（3）的 ρ of 比较依赖于内部表征。因此，它适用于具有椭圆等值集的非高斯噪声，也适用于每个试次包含更多刺激的测试，例如四刺激测试 [13], ，在该测试中，四个刺激（每种各两个）被分成两对。相同–不同测试的伯努利概率 [5, 6]也可以采用相同的变量约简。在相同–不同测试中，向受试者呈现两个刺激，并由其报告二者是否相同。内部表征为 $x \sim \mathcal{N}(\mu_{\mathrm{x}},\ \Sigma_{\mathrm{x}})$ 和 $y \sim \mathcal{N}(\mu_{\mathrm{y}},\ \Sigma_{\mathrm{y}})$；对于固定的 $\tau > 0$，当 $\rho(x,\ y) > \tau$ 时，模型认为受试者报告这两个刺激不同。

h 的网络近似有几个局限。其误差是在留出集上估算的，而非通过解析方法界定。网络的准确度取决于优化过程、训练点的数量及分布，以及训练标签的准确度。此外，针对每个维数 d 和误差容限 ϵ，都需要计算一个单独的网络（或特殊函数展开）。

致谢

我们感谢 Manas Rachh 在算法开发方面提出的许多有益讨论。部分软件开发、数值实验和论文编辑工作是在人工智能代理的协助下完成的。

参考文献

[1] Eric L. Amazeen. 通过动态触觉实现大小与重量的知觉独立性。《实验心理学杂志：人类感知与表现》，25(1):102– 119，1999年。[2] F. Gregory Ashby 和 James T. Townsend. 知觉独立性的多样性。*Psychological Review*，93(2):154–179，1986年。[3] Denise Benton 和 K. Krishnamoorthy. 计算连续分布的离散混合：非中心卡方、非中心 t 以及样本复相关系数平方的分布。

计算统计与数据分析，43(2):249–267，2003年。[4] James Bradbury、Roy Frostig、Peter Hawkins、Matthew James Johnson、Yash Katariya、Chris Leary、Dougal Maclaurin、George Necula、Adam Paszke、Jake VanderPlas、Skye Wanderman-Milne 和 Qiao Zhang. JAX：Python+NumPy 程序的可组合变换，2018年。http://github.com/jax-ml/jax。[5] Per B. Brockhoff 和 Rune Haubo Bojesen Christensen. 将瑟斯顿模型用于感觉辨别测试，作为广义线性模型。《食品质量与偏好》，21(3):330–338，2010年。[6] Rune Haubo Bojesen Christensen 和 Per B. Brockhoff. *sensR*——用于感觉辨别的 *R* 软件包，2023年。R 软件包版本 1.5-3。[7] 美国国家标准与技术研究院数学函数数字图书馆。https://dlmf.nist.gov/，2026年9月15日发布的 1.2.8 版。F. W. J. Olver、A. B. Olde Daalhuis、D. W. Lozier、B. I. Schneider、R. F. Boisvert、C. W. Clark、B. R. Miller、B. V. Saunders、H. S. Cohl 和 M. A. McClain 编。[8] Manfredo P. do Carmo. 黎曼几何。Birkh¨auser，波士顿，1992年。[9] Daniel M. Ennis. 感官辨别方法的效力。《感官研究杂志》，8(4):353–370，1993年。[10] Daniel M. Ennis 和 Norman L. Johnson. 用中心 F 分布函数表示偏好选择与三元选择的一般模型。*Psychometrika*，59(1):91–96，1994年。[11] Daniel M. Ennis 和 Kenneth Mullen. 维数对三角形方法所得结果的影响。《化学感官》，10(4):605–608，1985年。[12] Daniel M. Ennis 和 Kenneth Mullen. 辨别方法的多变量模型。《数学心理学杂志》，30:206–219，1986年。[13] John M. Ennis、Daniel M. Ennis、Dorene Yip 和 Michael O’Mahony. 四元组法变体的瑟斯顿模型。《英国数学与统计心理学杂志》，51(2):205–215，1998年。[14] 古斯塔夫·特奥多尔·费希纳。《心理物理学纲要》。Breitkopf und H¨artel，莱比锡，1860年。[15] 罗纳德·A·费希尔。《实验设计》。Oliver and Boyd，爱丁堡，1935年。[16] Daniel Fitousi 和 Michael J. Wenger. 面部身份与表情感知中的独立性变体。《实验心理学杂志：人类感知与表现》，39(1):133–155，2013年。[17] Wilson S. Geisler. 理想观察者理论对视觉研究的贡献。视觉研究，51(7):771–781，2011年。

[18] Amparo Gil、Javier Segura 和 Nico M. Temme。算法939：Marcum Q函数的计算。《*ACM*数学软件汇刊》，40(3):20:1–20:21，2014年。[19] David M. Green 和 John A. Swets。信号检测理论与心理物理学。Wiley，纽约，1966年。[20] Isabella Hartley、Liliana Orellana、Djin Gie Liem 和 Russell Keast。用于确定鲜味辨别状态的三角测试方法评估。《化学感官》，47:bjac003，2022年。[21] Matthew D. Hoffman 和 Andrew Gelman。无转向采样器：在哈密顿蒙特卡洛中自适应设定路径长度。《机器学习研究杂志》，15:1593–1623，2014年。[22] Fangfang Hong、Ruby Bouhassira、Jason Chow、Craig Sanders、Michael Shvartsman、Phillip Guan、Alex H Williams 和 David H Brainard。人类颜色辨别阈限的全面表征。*eLife*，14:RP108943，2026年7月。[23] Thomas Hummel、G. Kobal、H. Gudziol 和 A. Mackay-Sim。“嗅棒测试”的规范数据，包括气味识别、气味辨别和嗅觉阈限测试：基于3000多名受试者群体的更新。《欧洲耳鼻喉科学档案》，264(3):237–243，2007年。[24] Thomas Hummel、B. Sekinger、S. R. Wolf、E. Pauli 和 G. Kobal。“嗅棒测试”：通过结合气味识别、气味辨别和嗅觉阈限测试评估嗅觉表现。《化学感官》，22(1):39–52，1997年。[25] 国际标准化组织。感官分析——方法学——三角测试，2021年。ISO 4120:2021。[26] 国际标准化组织。感官分析——方法学——二-三点检验，2026年。ISO 10399:2026。[27] Norman L. Johnson、Samuel Kotz 和 N. Balakrishnan。连续单变量分布，第2卷。John Wiley & Sons，纽约，第2版，1995年。[28] John A. Kapenga、Kenneth Mullen、Elise de Doncker 和 Daniel M. Ennis。三角形方法中的多元正态密度函数积分。载于 Patrick Keast 和 Graeme Fairweather 编，《style id='32'>数值积分：最新进展、软件与应用，第203卷，北约高级研究所系列*C*：数学与物理科学，第321–328页。D. Reidel，多德雷赫特，1987年。[29] Min-A Kim、Young-Mi Lee 和 Hye-Seong Lee。比较 d' 用于辨别不同盐浓度番茄汁的二-三点检验三个版本所得的估计：参照刺激的数量和位置的影响。《食品质量与偏好》，21(5):504–511，2010年。[30] Diederik P. Kingma 和 Jimmy Ba。Adam：一种随机优化方法。载于第三届国际学习表征会议，*ICLR 2015*，2015年。

[31] David L. MacAdam。日光下对颜色差异的视觉敏感性。美国光学学会杂志，32(5):247–274，1942年。[32] Neil A. Macmillan 和 C. Douglas Creelman。检测理论：用户指南。Lawrence Erlbaum Associates，马瓦，新泽西州，第2版，2005年。[33] Jess I. Marcum。Q函数表。研究备忘录 RM-339，兰德公司，圣莫尼卡，加利福尼亚州，1950年。[34] Emma McMahon、Rozlynne Clarke、Rachael Jaenke 和 Julie Brimblecombe。在偏远澳大利亚原住民社区中检测面包中12.5%和25%的减盐幅度。*Nutrients*，8(3):169，2016年。[35] Takuma Morimoto。在真实世界照明环境中观察颜色与材料。在国际色觉研讨会上的报告，布莱顿，英国，2026年。2026年8月14日至18日。[36] Kenneth Mullen 和 Daniel M. Ennis。用于辨别方法的多元欧几里得模型的数学表述。*Psychometrika*，52(2):235–249，1987年。[37] Kenneth Mullen 和 Daniel M. Ennis。用于偏好选择和三元选择的简单多变量概率模型。*Psychometrika*，56(1):69–75，1991年。[38] Kenneth Mullen、Daniel M. Ennis、Elise de Doncker 和 John A. Kapenga。二-三点检验和三角形方法的模型。*Biometrics*，44(4):1169–1175，1988年。.[39] Robert Piessens、Elise de Doncker-Kapenga、Christoph W. Uberhuber 和 David K. Kahaner。*QUADPACK*：自动积分子程序包。Springer-Verlag，柏林，1983年。[40] Lorena Saavedra-Garcia、Antonio Bernabe-Ortiz、Robert H. Gilman、Francisco Diez-Canseco、Mar´ía Kathia C´ardenas、Katherine A. Sacksteder 和 J. Jaime Miranda。应用三角味觉测试评估低钠盐与普通食盐之间的差异：来自秘鲁的证据。*PLOS ONE*，10(7):e0134700，2015年。[41] Schematic。Super，2026年。Python软件包 schematic-supertest，版本0.1.1，https: //docs.schematic.tech/super，访问日期：2026-11-02。[42] Fabian A. Soto、Lauren Vucovich、Robert Musgrave 和 F. Gregory Ashby。具有个体差异的一般识别理论：一种研究知觉与决策交互作用的新方法，并应用于面孔知觉。*PsychonomicBulletin & Review*，22(1):88–111，2015年。[43] L. L. Thurstone。比较判断定律。*Psychological Review*，34(4):273–286，1927年。[44]Lloyd N. Trefethen。逼近理论与逼近实践。工业与应用数学学会，费城，2013年。[45] Pauli Virtanen、Ralf Gommers、Travis E. Oliphant 等。SciPy 1.0：用于科学计算的基础算法。*Nature Methods*，17(3):261–272，2020年。

A 理想观察者 ρ

用于三角测试的 [22] 模型通过距离函数比较内部表征 x_1, x_2 $\sim\mathcal{N}(\mu_{\mathrm{x}}$, $\Sigma_{\mathrm{x}})$ 和 $y \sim \mathcal{N}($ μ_{y}, $\Sigma_{\mathrm{y}})$ 之间的三个成对距离

$$\rho(x, y) = (x - y)^\top S^{-1}(x - y) \tag{57}$$

其中 $S = \frac{2}{3}\Sigma_{\mathrm{x}} + \frac{1}{3}\Sigma_{\mathrm{y}}$。因此，其决策规则假定已知内部表征 x_1, x_2, y和 Σ_{x}, Σ_{y}，但不知道真实刺激本身 μ_{x}，μ_{y}。在此处，在相同假设下，我们推导出理想观察者 [19, 17] 在三角测试和二选一三角测试任务中的决策规则。我们表明，这样的理想观察者使用相同的决策规则，只需将（57）中的 S 替换为 $S = \Sigma_{\mathrm{x}}$。真实观察者可能无法获取Σ_{x}, Σ_{y}，但在呈现了 μ_{x}，μ_{y}附近足够多的刺激后，观察者或许能够合理地估计 Σ_{x}, Σ_{y}。理想观察者已被证明在视觉研究中非常有用，既可作为衡量人类表现的基准，也可作为真实观察者的原理性模型 [17]。

我们首先假设，已知 $\Sigma_{\mathrm{x}}, \Sigma_{\mathrm{y}}$ 的理想观察者观测到内部表征 w_0, w_1, w_2，其中 w_i, $w_j \sim \mathcal{N}(\mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})$ $w_k \sim \mathcal{N}(\mu_{\mathrm{y}}, \Sigma_{\mathrm{y}})$，对于某个 i, j, k，使得 $\{i, j, k\} = \{0, 1, 2\}$。我们假设理想观察者不知道哪个是独特刺激。也就是说，理想观察者不知道 $k \in \{0, 1, 2\}$ 以及 $\{i, j\}$ $=\{0, 1, 2\} \setminus \{k\}$。

在我们的模型下（我们假设理想观察者知道该模型），有三个未知参数—— $\mu_{\mathrm{x}}, \mu_{\mathrm{y}}$, 以及独特刺激的索引 $k \in \{0, 1, 2\}$。对于任意 $k \in \{0, 1, 2\}$和 $\{i, j\} = \{0, 1, 2\} \setminus \{k\}$, w_k是独特刺激的内部表征的似然为

$$p(w_0, w_1, w_2 \mid k, \mu_{\mathrm{x}}, \mu_{\mathrm{y}}) = \phi(w_i; \mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})\, \phi(w_j; \mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})\, \phi(w_k; \mu_{\mathrm{y}}, \Sigma_{\mathrm{y}}). \tag{58}$$

理想观察者对 $k \in \{0, 1, 2\}$ 采用均匀先验，对μ_{x}采用均匀（非正规）先验，μ_{y} ，并选择

$$\hat{k}^{\mathrm{tri}} := \arg\max_{\kappa} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} p(w_0, w_1, w_2 \mid \kappa, \mu_{\mathrm{x}}, \mu_{\mathrm{y}})\, d\mu_{\mathrm{y}}\, d\mu_{\mathrm{x}}. \tag{59}$$

尽管对 $\mathbb{R}^d$ 采用的均匀先验是非正规的，但积分（59）的被积函数是高斯分布，因此积分有限。根据理想观察者已知的信息，能使识别出独特刺激的后验概率达到最大的刺激选择是 $\hat{}k^{\mathrm{tri}}$。

在二-三点检验中，我们有一个参照刺激，其内部表征记为 $w_0 \sim \mathcal{N}(\mu_{\mathrm{x}}$, $\Sigma_{\mathrm{x}})$，另外还有两个刺激，其内部表征为 w_1, w_2，其中w_k $\sim \mathcal{N}(\mu_{\mathrm{x}}$, $\Sigma_{\mathrm{x}})$、 $w_{\sim k}$ $\sim \mathcal{N}(\mu_{\mathrm{y}}$, $\Sigma_{\mathrm{y}})$，对于 $\{k, {\sim}k\} = \{1, 2\}$。 w_k是参照刺激重复样本的内部表征的似然为

$$p(w_1, w_2 \mid w_0, k, \mu_{\mathrm{x}}, \mu_{\mathrm{y}}) = \phi(w_0; \mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})\, \phi(w_k; \mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})\, \phi(w_{\sim k}; \mu_{\mathrm{y}}, \Sigma_{\mathrm{y}}), \tag{60}$$

而对 k, μ_{x},μ_{y}采用均匀先验的理想观察者会选择

$$\hat{k}^{\mathrm{duo}} := \arg\max_{\kappa \in \{1,2\}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(w_0; \mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})\, \phi(w_\kappa; \mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})\, \phi(w_{\sim\kappa}; \mu_{\mathrm{y}}, \Sigma_{\mathrm{y}})\, d\mu_{\mathrm{y}}\, d\mu_{\mathrm{x}}. \tag{61}$$

定义。 三角测试中理想观察者的响应为$\hat{}k^{\mathrm{tri}}$ （*59*），二-三点检验中的响应为$\hat{}k^{\mathrm{duo}}$ （61）。

下面的结果是本节的主要结果，它表明，$\hat{}k^{\mathrm{tri}}$ 和$\hat{}k^{\mathrm{duo}}$ 正是规则 (1) 和 (2) 在给定 $S = \Sigma_{\mathrm{x}}$ 时所作出的选择，因此 p_{tri} 和 p_{duo} 是理想观察者作出正确回答的概率。

命题 4。 对于所有 w_0、w_1、$w_2 \in \mathbb{R}^d$，理想观察者为三角测试所选出的独特刺激 $\hat{}k^{\mathrm{tri}}$（见 (59)）满足

$$\hat{k}^{\mathrm{tri}} = \operatorname*{argmin}_{\kappa\in\{0,1,2\}} (w_i - w_j)^\top \Sigma_{\mathrm{x}}^{-1}(w_i - w_j) \tag{62}$$

其中 $\kappa \in \{0,1,2\}$和 $\{i,j\} = \{0,1,2\} \setminus \{\kappa\}$。二-三点检验中的选择 $\hat{}k^{\mathrm{duo}}$ （见 (61)）满足

$$\hat{k}^{\mathrm{duo}} = \operatorname*{argmin}_{\kappa\in\{1,2\}} (w_0 - w_\kappa)^\top \Sigma_{\mathrm{x}}^{-1}(w_0 - w_\kappa). \tag{63}$$

此外，对于形式为 (3) 且满足 $S = \Sigma_{\mathrm{x}}$ 的 ρ，

$$p_{\mathrm{tri}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma_{\mathrm{x}}, \Sigma_{\mathrm{y}}) = \Pr(\hat{k}^{\mathrm{tri}} = k), \tag{64}$$

$$p_{\mathrm{duo}}(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}, \Sigma_{\mathrm{x}}, \Sigma_{\mathrm{y}}) = \Pr(\hat{k}^{\mathrm{duo}} = k). \tag{65}$$

证明。 固定 w_0、w_1、$w_2 \in \mathbb{R}^d$和 $\kappa \in \{0,1,2\}$，并令 $\{i,j\} = \{0,1,2\} \setminus \{\kappa\}$。根据高斯概率密度函数的定义 ϕ，可立即得到

$$\int_{\mathbb{R}^d} \phi(w_\kappa; \mu_{\mathrm{y}}, \Sigma_{\mathrm{y}})\, d\mu_{\mathrm{y}} = \int_{\mathbb{R}^d} \phi(\mu_{\mathrm{y}}; w_\kappa, \Sigma_{\mathrm{y}})\, d\mu_{\mathrm{y}} = 1. \tag{66}$$

将（66）和（58）代入（59），得到

$$\hat{k}^{\mathrm{tri}} = \arg\max_{\kappa} \int_{\mathbb{R}^d} \phi(w_i; \mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})\, \phi(w_j; \mu_{\mathrm{x}}, \Sigma_{\mathrm{x}})\; d\mu_{\mathrm{x}}. \tag{67}$$

展开高斯概率密度函数 ϕ，并注意到归一化常数与 κ无关，得到

$$\hat{k}^{\mathrm{tri}} = \arg\max_{\kappa} \int_{\mathbb{R}^d} \exp\left(-\tfrac{1}{2}\left[(w_i - \mu_{\mathrm{x}})^\top \Sigma_{\mathrm{x}}^{-1}(w_i - \mu_{\mathrm{x}}) + (w_j - \mu_{\mathrm{x}})^\top \Sigma_{\mathrm{x}}^{-1}(w_j - \mu_{\mathrm{x}})\right]\right) d\mu_{\mathrm{x}} \tag{68}$$

我们将高斯被积函数拆分为两个因子的乘积：一个高斯分布依赖于 μ_{x}，另一个则独立于 μ_{x},

$$\hat{k}^{\mathrm{tri}} = \arg\max_{\kappa} \int_{\mathbb{R}^d} \exp\Big(-\tfrac{1}{2}\left[(w_i - \mu_{\mathrm{x}})^\top \Sigma_{\mathrm{x}}^{-1}(w_i - \mu_{\mathrm{x}}) + (w_j - \mu_{\mathrm{x}})^\top \Sigma_{\mathrm{x}}^{-1}(w_j - \mu_{\mathrm{x}})\right]\Big) d\mu_{\mathrm{x}} \tag{69}$$

$$= \arg\max_{\kappa} \int_{\mathbb{R}^d} \exp\Big(-\tfrac{1}{4}(w_i - w_j)^\top \Sigma_{\mathrm{x}}^{-1}(w_i - w_j)\; - (\mu_{\mathrm{x}} - \alpha)^\top \Sigma_{\mathrm{x}}^{-1}(\mu_{\mathrm{x}} - \alpha)\Big) d\mu_{\mathrm{x}} \tag{70}$$

$$= \arg\max_{\kappa}\; \exp\Big(-\tfrac{1}{4}(w_i - w_j)^\top \Sigma_{\mathrm{x}}^{-1}(w_i - w_j)\Big) \int_{\mathbb{R}^d} \exp\Big(-(\mu_{\mathrm{x}} - \alpha)^\top \Sigma_{\mathrm{x}}^{-1}(\mu_{\mathrm{x}} - \alpha)\Big) d\mu_{\mathrm{x}} \tag{71}$$

$$= \arg\max_{\kappa}\; \exp\Big(-\tfrac{1}{4}(w_i - w_j)^\top \Sigma_{\mathrm{x}}^{-1}(w_i - w_j)\Big) \tag{72}$$

其中 $\alpha = (w_i + w_j)/2$。

对于二-三点检验，将 (w_i, w_j, w_κ) 替换为 $(w_0, w_\kappa, w_{\sim\kappa})$，进行相同的分析，可得 (61) 中的积分等于 $\det(4\pi\Sigma_{\mathrm{x}})^{-1/2}\exp\big(-\frac{1}{4}(w_0 - w_\kappa)^\top \Sigma_{\mathrm{x}}^{-1}(w_0 - w_\kappa)\big)$，由此得到 (63)。

对于 (64)，令 $(w_i, w_j, w_k) = (x_1, x_2, y)$。根据 (62)，$\hat{k}^{\mathrm{tri}} = k$ 当且仅当 $\rho(x_1, x_2) < \min\{\rho(x_1, y)\ \rho(x_2, y)\}$，这与 (1) 中的事件相同。同理，令 $(w_0, w_k, w_{\sim k}) = (x_{\mathrm{ref}}, x_1, y_1)$，由 (63) 可得 $\hat{k}^{\mathrm{duo}} = k$ 当且仅当 $\rho(x_{\mathrm{ref}}, x_1) < \rho(x_{\mathrm{ref}}, y_1)$，即 (2) 中的事件。 □

备注4. 当(59)中的积分替换为对 μ_{x}、μ_{y}取最大值时，(62)中的$\hat{k}^{\mathrm{tri}}$ 的选择保持不变。二-三点检验也同样如此。事实上，对于每个 κ,

$$\begin{aligned}&\max_{\mu_{\mathrm{x}},\mu_{\mathrm{y}}} p(w_0, w_1, w_2 \mid \kappa, \mu_{\mathrm{x}}, \mu_{\mathrm{y}}) \\ &\qquad = \det(\pi\Sigma_{\mathrm{x}})^{-1/2}\det(2\pi\Sigma_{\mathrm{y}})^{-1/2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} p(w_0, w_1, w_2 \mid \kappa, \mu_{\mathrm{x}}, \mu_{\mathrm{y}})\, d\mu_{\mathrm{y}}\, d\mu_{\mathrm{x}},\end{aligned}$$

其中常数不依赖于 κ。