arXiv:2611.30822v1 [stat.ML] 23 Nov 2026

图模型中的快速且一致的结构学习：通过近似交叉验证

Zhao Lyu
Univers ity of Ch
芝加哥大学
zlyu@uchicago.edu

Mladen Kolar南加州大学 & 穆罕默德·本·扎耶德人工智能大学
mkolar@marshall.usc.edu

Bryon Aragam
University of Chi
芝加哥
bryon@chicagobooth.edu

摘要

高维稀疏图模型的结构学习，既要求具备可扩展的计算能力，也要求满足选择一致性。贝叶斯信息准则及扩展贝叶斯信息准则等经典准则可以保证一致性，但依赖于特定模型的似然和复杂度校准。相比之下，交叉验证无需进行此类解析校准，因此在图学习中颇具吸引力。然而，交叉验证在结构学习中的计算开销难以承受，而且在理论上并不一致。这使得一致性与建模灵活性之间产生了矛盾。在本研究中，我们通过开发一种基于近似交叉验证、可扩展的图恢复方法，化解了这一矛盾。我们从理论和实证两方面展示了该方法在学习大型图模型方面的有效性，并证明在较弱的信号强度假设下，该方法具有选择一致性。

1 引言

高维结构学习旨在恢复潜在结构，而不仅仅是优化预测度量。在图模型中，这种结构由边集编码。因此，核心挑战是一个选择问题：面对众多候选结构，我们需要一种兼具计算可行性和统计正确性的规则。

贝叶斯信息准则和扩展贝叶斯信息准则（BIC [38], EBIC [11]）等基于似然校准的经典准则可以提供一致性保证，但需要特定模型的似然，并通过惩罚项进行复杂度校准。交叉验证（CV）提供了另一种途径 [1, 3, 19, 41]：它不依赖特定模型，也就是说，一旦指定验证损失，同一经验规则便可比较候选拟合方法，而无需推导专门的惩罚项。它还基于经验进行校准：它在留出样本上评估候选方法，而非依赖针对特定模型类别推导出的固定解析惩罚项。这些特性使交叉验证适用于结构学习，但精确交叉验证需要反复重新拟合，对于大型图模型而言，计算开销难以承受。

近似交叉验证（ACV）通过以低成本近似替代重复重新拟合，直接解决这一计算瓶颈 [4, 17, 34, 43]。现有理论主要关注这些近似能否保留验证损失或模型评估质量 [39, 40, 46]。然而，对于结构学习而言，仅近似损失并不足够：候选模型可能在留出数据上表现良好，却同时选错结构。因此，关键问题不在于 ACV 是否能像交叉验证那样估计留出损失，而在于能否从其选出的结构中恢复真实结构。

在本文中，我们对 ACV 作为结构学习规则进行了全面刻画，并从结构层面回答了这些问题。ACV 本身并不是一种一致的结构学习规则，因为它继承了交叉验证的过度选择行为。然而，这一缺陷最终证明是有用的，而非致命的。对过度选择的一种不那么悲观的解读是：它仍然能够提供高质量的初步筛选。我们证明，ACV 选择的结构以较高概率包含真实结构，

第40届神经信息处理系统大会（NeurIPS 2026）。

从而使 ACV 成为一种可扩展的图恢复中间估计。自然地，在此基础上进行适当的后筛选步骤，我们便可以移除额外边，并实现选择一致性。我们的主要贡献总结如下：

1. **负面结果：不一致性**。我们分析了为何仅使用 ACV 无法作为最终的结构学习规则：由于过度选择，它继承了交叉验证的选择不一致性。2. **正面结果：后筛选一致性**。我们从互补的角度看待 ACV 的失败，并将 ACV 视为一种可扩展的筛选方法。在中间 ACV 估计的基础上进行适当的后校正，可以移除额外边，从而实现选择一致性。3. **实证验证**。我们通过实证展示了计算-统计权衡：ACV 提供快速且高召回率的筛选，而校正后的方法则改善了最终的图恢复效果。

1.1 问题设定

令 subG(σ^2)[1] 表示 $\mathbb{R}^p$ 上参数为 σ^2的零均值次高斯分布族。固定 $P \in \text{subG}(\sigma^2)$，并令 $X = [X_1, X_2, \ldots, X_p] \sim P$。 X 的条件独立结构由一个马尔可夫随机场表示，即一个无向图 $G = (V, E)$，其节点集为 $V = [p]$。具体而言，如果节点 i 与节点 j之间不存在边，则在给定其余变量$X_{\setminus\{i,j\}}$ 的条件下， X_i 与 X_j 条件独立。这就是局部马尔可夫性质：每个变量在给定其图邻居的条件下，与其所有非邻居条件独立。

我们考虑根据 n 个来自 X 的独立同分布观测值，学习 G 的结构，数据由数据矩阵 $\mathbf{X}$ 表示。组合后的数据矩阵 $\mathbf{X}$ 的每一行向量都是从 P 抽取的独立同分布样本，且$\Sigma = \text{Cov}(X) \succ 0$ 表示总体协方差。若 $X \sim \mathcal{N}(\mathbf{0}, \Sigma)$，精度矩阵 Σ^{-1}通过其零模式编码图的边结构：当且仅当 $(i, j) \notin E$ 时， $\Theta_{ij} = 0$。这定义了一个关于 X 的有效马尔可夫随机场 G，而恢复 G 可归结为估计 Σ^{-1} 的零模式。许多方法都用于估计精度矩阵以进行结构学习，例如基于乔列斯基因子的方法 [5, 15, 47] 和惩罚似然方法 [8, 12, 37, 49]。特别地，对于任意矩阵 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 和 $\lambda > 0$，常用的 ℓ_1 惩罚负对数似然估计量 [2, 12, 36]定义为：

$$\widehat{\Theta}^{\lambda,\mathbf{X}} := \arg\min_{\Theta\succ 0} -\log\det\Theta + \text{Tr}(S(\mathbf{X})\Theta) + \lambda\|\Theta\|_1, \tag{1}$$

其中， $S(\mathbf{X}) = \mathbf{X}^\top\mathbf{X}/n$ 是依赖于$\mathbf{X}$[2], 和 $\|\Theta\|_1 = \sum_{i,j}|\Theta_{ij}|$的经验协方差矩阵。估计图（零模式）定义为

$$\widehat{G}^{\lambda,\mathbf{X}} := \text{supp}(\widehat{\Theta}^{\lambda,\mathbf{X}}) = \{(i,j) : \widehat{\Theta}^{\lambda,\mathbf{X}}_{ij} \neq 0\}. \tag{2}$$

值得注意的是，精度矩阵的严格解释仅适用于通常的高斯图模型设定及相关的半参数扩展。对于非高斯数据，Σ^{-1} 的零模式未必能在没有额外假设的情况下表示条件独立性，但基于精度矩阵的估计量仍被广泛用于恢复结构关系：例如，基于 copula 的方法 [23], 基于秩的方法 [10, 48]以及稳健惩罚项 [26]。

在本文中，我们通过式（1）中的精度矩阵估计量的支撑集研究稀疏无向图模型选择。采用次高斯性是为了简化集中性论证；第 6.3 节以及附录 E.3 和 E.4 中的稳健性实验和压力测试表明，我们的方法适用于次高斯假设之外的情形。

1.2 图选择的交叉验证

在稀疏图模型中，选择 λ 等价于选择式（2）中的图 $\widehat{G}^{\lambda,\ \mathbf{X}}$。因此，交叉验证直接决定了学习得到的图。具体而言，留一交叉验证每次留出一个样本，以构造 n Θ 的估计，并评估该估计在这个留出

[1]若存在 $\sigma \in (0, \infty)$ 使得 $\mathrm{E}[e^{tZ}] \le e^{t^2\sigma^2/2}\ \forall t \in \mathbb{R}$，则称零均值随机变量 Z 是次高斯的。这里，subG(σ^2) 由满足以下条件的零均值分布 P 构成：对于具有协方差 Σ 的 $X \sim P$，每个 $X_i/\sqrt{\Sigma_{ii}}$ 都满足上述次高斯界，其中 σ。[2]$S(Y) = YY^\top$ 表示单个列向量 Y。在表示 $S(\mathbf{X}^{(i)})$ 时，我们将第 i 行 $\mathbf{X}^{(i)}$ 作为列向量使用。

样本上的表现。对于 $i \in [n]$，我们将 $\mathbf{X}^{(-i)}$（训练时排除其第 i 行的 $\mathbf{X}$ 的（$n-1$）×p 子矩阵）和 $\mathbf{X}^{(i)}$（验证时排除 $\mathbf{X}$的第 i 行的（$n-1$）×p 子矩阵）分别记为。选定图为 $\widehat{G}^{\lambda^{\mathbf{X}}}{}_{\text{交叉验证}}{}_{,\mathbf{X}}$，其中选定的惩罚项定义如下:

$$\lambda_{\mathrm{CV}}^{\mathbf{X}} := \arg\min_{\lambda>0} \frac{1}{n}\sum_{i=1}^{n} -\log\det\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} + \mathsf{Tr}\left(S(\mathbf{X}^{(i)})\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\right).$$

作为技术分析中的一种工具，我们还定义了预言机惩罚项选择 $\lambda_*^{\mathbf{X}}$:

$$\lambda_*^{\mathbf{X}} := \arg\min_{\lambda>0} \mathsf{E}_{Y\sim P} -\log\det\widehat{\Theta}^{\lambda,\mathbf{X}} + \mathsf{Tr}\left(S(Y)\widehat{\Theta}^{\lambda,\mathbf{X}}\right). \tag{3}$$

尽管不同的划分策略在计算成本与准确性之间实现了权衡，但 CV 的核心成本在于反复重新拟合。许多现有研究通过从热启动构造留一估计来降低这一成本。例如，利用全数据估计量执行牛顿步，以近似基于训练集 [4, 34, 39, 40, 46] 的优化目标。这些近似为获得交叉验证选定图提供了一条计算途径，但统计问题仍然存在：所得图能否恢复正确的结构?

我们使用 ACV 来应对计算障碍，随后分析由该近似选定的图。我们重点关注留一交叉验证，因为与其他 CV 变体相比，它能够更准确地评估模型性能 [1, 34]；在大规模实验中则使用 a K-折版本。

本文其余部分的组织如下。我们在第 2 节定义 ACV 选定的图。第 3 节和第 4 节说明了为何仅使用 ACV 无法实现一致性，尽管它仍可作为筛选后的候选图。在第 5 节中，我们提出一种能够恢复选择一致性的校正步骤；第 6 节通过实证评估由此产生的计算-统计权衡，并验证我们用于受控恢复的流程。

2 用于图选择的近似交叉验证

令 $\widehat{\mathcal{S}} := \mathrm{supp}(\widehat{\Theta}^{\lambda,\ \mathbf{X}})$ 为 $\lambda > 0$ 的全数据估计量的支撑集。我们以全数据估计量作为热启动，构造留一交叉验证估计量 $\widehat{\Theta}^{\lambda,\ \mathbf{X}^{(-i)}}$ 的近似值 $\widetilde{\Theta}^{\lambda,\ \mathbf{X}^{(-i)}}$：

$$\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} := \begin{bmatrix} \widehat{\Theta}^{\lambda,\mathbf{X}}_{\widehat{\mathcal{S}}} + \left(\widehat{\Theta}^{\lambda,\mathbf{X}} \otimes \widehat{\Theta}^{\lambda,\mathbf{X}}\right)_{\widehat{\mathcal{S}}\widehat{\mathcal{S}}} \left(S(\mathbf{X}) - S(\mathbf{X}^{(-i)})\right)_{\widehat{\mathcal{S}}} \\ \mathbf{0}_{\setminus\widehat{\mathcal{S}}} \end{bmatrix}, \tag{4}$$

其中，$\otimes$ 表示克罗内克积，坐标限制按向量化方式解释。我们使用全数据优化器作为热启动，并在限制条件之间应用光滑似然˜项的海森矩阵逆，以求得 $\mathcal{S}$。支撑限制的依据是针对非光滑惩罚项的 ACV 构造 [39]；附录 A.1 直接分析了所示更新。

尽管 $\widehat{\Theta}^{\lambda,\ \mathbf{X}}$ 的正定性使得能够显式计算海森矩阵逆 [39], ，但矩阵运算的规模会随克罗内克限制支撑集而增大。在我们的实现中，我们通过向量化样本协方差差而非构造完整的克罗内克积来利用稀疏性。所有留一交叉验证的近似方法仍会随留出样本的数量而扩展，因此对于较大的 n，我们采用类似的 K 折 ACV 评分，其留出拟合的数量由 K [30] 固定。经 ACV 调优的惩罚项 $\lambda_{\mathrm{ACV}}^{\mathbf{X}}$ 是使验证损失最小的惩罚项:

$$\lambda_{\mathrm{ACV}}^{\mathbf{X}} := \arg\min_{\lambda>0} \frac{1}{n}\sum_{i=1}^{n} \left[-\log\det\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} + \mathsf{Tr}\Big(S(\mathbf{X}^{(i)})\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\Big)\right]. \tag{5}$$

当其参数不是正定矩阵时，对数行列式被视为无效，并赋值为 $+\infty$。若 ACV 下确界为 $-\infty$，则设 $\lambda_{\mathrm{ACV}}^{\mathbf{X}} = 1 + \max_j S(\mathbf{X})_{jj}$。在 $\lambda_*^{\mathbf{X}}$和 $\lambda_{\mathrm{ACV}}^{\mathbf{X}}$的定义中，argmin 表示在最小值可达时的可测精确正极小值点；否则，表示一个可测正值点，其目标函数值与有限下确界相差不超过 n^{-2}。相关存在性和细节将在附录 A.3 中讨论。于是，ACV 选定的图为 $\widehat{G}^{\lambda^{\mathbf{X}}}{}_{\mathrm{ACV}}{}_{,\mathbf{X}}$。

3 ACV 的不一致性

根据构造，ACV 旨在近似交叉验证估计，并自然地继承了结构不一致性。为更好地理解二者之间的联系，我们首先证明，式 (3) 中定义的预言机惩罚项可以通过式 (5) 中经 ACV 调整的 λ 得到良好估计。

引理 3.1（与预言机的联系）。 对于每个 $\delta > 0$,

$$\mathsf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(|\lambda^{\mathbf{X}}_{\mathrm{ACV}} - \lambda^{\mathbf{X}}_{*}| < \delta) \to 1 \quad as\ n \to \infty.$$

当出于计算量方面的考虑而用 ACV 替代交叉验证时， λ_{ACV} 的渐近行为很有用。通过将 λ_{ACV} 与 λ_{*}联系起来，我们能够从预言机 λ_{*}的角度评估 ACV 在结构学习中的表现。

考虑高斯模型，在适当调整 λ [36, 42]的情况下，可以实现一致的图恢复。

Theo引理 3.2 （ACV 不一致性）。 假设 $p \geq 2$和 $\Sigma \succ 0$ 固定，且每一行 $\mathbf{X}$是一个来自 $\mathcal{N}(0, \Sigma)$ 的独立样本。如果关联图 G 不是完全图，那么

$$\limsup_{n\to\infty} \mathbf{Pr}_{\mathbf{X}\sim N(0,\Sigma)^{\otimes n}} \left(\widehat{G}^{\lambda^{\mathbf{X}}_{\mathrm{ACV}},\mathbf{X}} = G\right) < 1.$$

定理 3.2 表明，对于特殊的高斯图模型，基于 ACV 选择 λ 在结构学习中可被证明是不一致的。这与已有研究中关于通过交叉验证选择调参来进行结构学习会导致不一致的结果相符，即使对于足够稀疏的图 [28]也是如此。这与已有研究展示的结果形成对比，例如 ACV 具有风险一致性。换言之，尽管存在经过适当调参即可实现一致性的 λ，但正是 ACV 对 λ 的错误选择导致了结构学习中的不一致性。这种结构不一致性也与风险一致性并存。因此，我们需要探究这种现象发生的原因、机制，以及如何纠正。

4 预筛选：关于不一致性的另一种视角

为了解释选择不一致性，需要考虑三种情况。两个极端情况分别是漏掉真实边，或错误地纳入不正确的边。第三种情况介于两者之间，既漏掉真实边，也纳入虚假边。根据引理 3.1，值得考察一下预言机。

考虑 $\Theta^{\lambda_*,\ \mathbf{X}}$ 这个在风险意义上达到最佳模型性能的预言机模型。遗漏任何特征都可能降低其性能，这与定义（3）相矛盾。另一方面，良好的预测性能表明某种 具有信息量的结构已经被学习到，因此或许可以恢复一致性。这一直觉意味着必须恢复所有真实边，从而排除遗漏真实边的可能性。这为理解不一致性提供了一个更有启发性的视角。

定理 4.1 （预筛选）。假设 $\Sigma \succ 0$ 和（*A3*）成立。那么， $\widehat{G}^{\lambda^{\mathbf{X}}\ \mathrm{ACV},\mathbf{X}}$ 在（2）中定义，具有预筛选性质：它以高概率选出每一条真实边，即

$$\mathsf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(G \subseteq \widehat{G}^{\lambda^{\mathbf{X}}_{\mathrm{ACV}},\mathbf{X}}) \to 1\ as\ n \to \infty.$$

定理 4.1 可由引理 3.1 证明中建立的总体风险界推出；完整证明见附录 B.1。回顾一下，定理 3.2 证明了 ACV 选定图的不一致性，而定理 4.1 证明了预筛选性质。这些结果共同表明，剩余差异是额外边，而非遗漏的真实边。我们不将此视为负面结果，而是换个角度看待，将不一致性转化为一个机会：如果能成功滤除虚假边，就可以恢复一致性。

5 校正 ACV 选定的图

我们所关注的量是图估计的选择一致性，即估计图能否恢复真实的图结构。定理 4.1 中的预筛选性质提出了一种筛选

方案，以排除 $\widehat{G}^{\lambda^{\mathbf{X}}}{}_{\text{ACV},\mathbf{X}}$ 中错误纳入的边，而可区分性通常依赖于信号强度。为使模型选择问题具有可识别性，我们作出以下一般性假设，以明确高维图结构恢复的设置。

(A1) 高维性。 存在 $\tau > 0$，使得维度 p 随 n 增长，其中 $p \le C_1 n^{\tau}$，且某个常数 $C_1 > 0$。(A2) 非奇异性。 协方差矩阵为正定，$\mathrm{Cov}(X) \succ 0$，且条件数 $\kappa(\Sigma) \le C_2$ 有界，其中某个常数 $C_2 > 0$。(A3) 信号强度。 存在某个常数 $C_3 > 0$ 和 $0 \le \gamma < \frac{1}{2}$，使得对于每个 $(i,j) \in E$，

$$|\rho_{ij|G}| \ge C_3 n^{-\gamma},$$

其中 $\rho_{ij|G} = -\frac{(\Sigma^{-1})_{ij}}{\sqrt{(\Sigma^{-1})_{ii}(\Sigma^{-1})_{jj}}}$ 是全阶线性偏相关。

(A4) 线性。 令 $\mathrm{ne}(j)$ 表示图 G 中节点 j 的邻居。对于每个 $j \in V$，给定其图邻居时，X_j 的条件均值是线性的：$X_j = \theta_j^{\top} X_{\mathrm{ne}(j)} + \varepsilon_j$，其中 $\varepsilon_j \perp X_{\mathrm{ne}(j)}$ 且 $\mathrm{E}[\varepsilon_j \mid X_{\mathrm{ne}(j)}] = 0$。

需要指出的是，(A1)-(A3) 是高维恢复中的标准正则条件。(A4) 是一项简化假设，使理论目标限定在稀疏精度图恢复的范围内；分析一般的非线性条件独立图则需要采用不同的非参数假设，这超出了本文的研究范围。E.4 节中的稳健性实验评估了该过程在这一形式化设定之外的表现。

在这些假设下，我们应用算法 1 中的筛选后方案，并得出校正后的 ACV 估计具有一致性。

定理 5.1. 设 $\check{G}^{\lambda^{\mathbf{X}}}{}_{\text{ACV},\mathbf{X}}$ 表示算法 *1* 在输入为 $\widehat{G} = \widehat{G}^{\lambda^{\mathbf{X}}}{}_{\text{ACV},\mathbf{X}}$ 和 $\alpha = n^{-\varepsilon}$、固定 $\varepsilon > 2\tau$ 时的输出。那么，在假设 *(A1)-(A4)* 下

$$\mathrm{Pr}_{\mathbf{X}\sim P^{\otimes n}}(\check{G}^{\lambda^{\mathbf{X}}_{\mathrm{ACV}}} = G) \to 1 \; \textit{as} \; n \to \infty.$$

定理 5.1 表明，额外的后校正步骤能够获得具有λ_{ACV}的一致图估计。我们强调，$\check{G}^{\lambda_{\mathrm{ACV}},\mathbf{X}}$ 的一致性并不与已有的不一致性结果相矛盾。显然，交叉验证在一个步骤中处理并选择模型，而 $\check{G}$ 则分两个步骤进行图估计。换言之，ACV 通过预筛选准备充分的信息，以便在第二阶段（后校正）中进一步细化。

本质上，算法 1 构造了一个归一化偏相关统计量 T，用于检验真实偏相关是否为零。此处 $\mathrm{Cov}(e_i,\ e_j) = n^{-1}\sum_{k=1}^{n} e_i^{(k)} e_j^{(k)}$。最小二乘残差通过正交投影定义；在实际应用中，如果其分母为零，我们就设定 $T_{ij} = 0$（详情参见附录 D.2）。归一化统计量的一般思路与许多阈值化方法 [2, 6, 27, 29]相似。在零假设下，该归一化统计量在检验阈值所用范围内满足一致的正态尾部近似。我们在以下定理 5.2 中证明了该算法的有效性：

定理 5.2。 假设我们以满足 $G \supseteq G$ 的 G 初始化算法 *1*。在假设 *(A1)*、*(A2)* 和 *(A4)* 下，对于 $\alpha \in (0,\ 1)$

$$\mathrm{Pr}_{\mathbf{X}\sim P^{\otimes n}}(\check{G}^{\alpha,\mathbf{X}} \not\subseteq G) \le \alpha \quad \textit{as} \; n \to \infty,$$

其中，$\check{G}^{\alpha,\ \mathbf{X}}$ 是通过算法 *1* 构造的。

值得注意的是，由于定理 4.1 中的预筛选性质，该要求不会对 ACV 及其他 CV 变体造成任何限制。对于大型稀疏图，直接应用于整个图会使计算负担落在独立性检验上，从而浪费 ACV 在可扩展性和效率方面的优势，而这些优势本可用于缩小估计边集。还需指出，后校正步骤可通过其他方案广泛地实现自适应，例如排序和扰动检验。这里的关键在于，结构信息确实可以从 ACV 中恢复，因此能够采用高效、自适应且与模型无关的方法，提取具有保证的图估计。

算法 1 针对 ACV 调优后的 λ_{ACV} 进行事后校正

要求： 估计图 $G=(V,\ E)$，其 $G\subseteq G$，水平 $\alpha\in(0,\ 1)$，数据矩阵 $\mathbf{X}\ \in\mathbb{R}^{n\times p}$**确保：** 精化后的图估计 $\check{G}^{\alpha,\ \mathbf{X}}$

1：**对**每条边 $(i,\ j)\ \in E$ **执行**

2：求出 G中邻居的并集：

$$U(i,j)=\{k\in[p]\setminus\{i,j\}:(i,k)\in\widehat{E}\vee(j,k)\in\widehat{E}\}$$

3：通过对$\mathbf{X}_i$（相应地，对$\mathbf{X}j$）关于$\mathbf{X}^{U(i,j)}$进行普通最小二乘回归，计算残差 e_i（相应地，e_j）4：按如下方式构造标准化偏相关检验统计量

$$\widehat{T}_{ij}=\sqrt{\frac{n}{\frac{1}{n}\sum_{k=1}^{n}\left(e_i^{(k)}e_j^{(k)}-\mathsf{Cov}(e_i,e_j)\right)^2}}\mathsf{Cov}(e_i,e_j)$$

5：构造精化后的边集和图估计，其定义如下

$$\check{E}^{\alpha,\mathbf{X}}=\left\{(i,j)\in\widehat{E}:|\widehat{T}_{ij}|\geq\Phi^{-1}\left(1-\frac{\alpha}{2p^2}\right)\right\},$$
$$\check{G}^{\alpha,\mathbf{X}}=(V,\check{E}^{\alpha,\mathbf{X}}),$$

其中，Φ 表示标准正态分布的累积分布函数。

6 实验

为进行演示，我们在 k 近邻图上进行所有比较，数据由不同分布生成，并使用 $k=1$。真实图通过将每个顶点与其k个最近邻相连构造而成，邻居关系依据随机生成的度量确定。不同图类型对应不同的稀疏性，详细的图构造方法见附录 E.1。

我们总共考虑四种方法：结合算法 1 中后校正的 ACV、K折交叉验证（CV K折）、贝叶斯信息准则和扩展贝叶斯信息准则。除非另有说明，ACV 和 CV 均使用 $K=3$，其他 CV 变体的比较则留待附录讨论。前两种是与模型无关的方法，后两种则是经典的一致性方法。我们还在附录 E.6 中加入与稳定性选择 [24], TIGER [22], 和 Lockhart协方差检验 [25]的比较。为从不同角度考察稀疏正则化方法，我们采用以下数据集设置：线性高斯、线性非高斯、非线性、弱信号以及真实基因数据集。我们分别展示预筛选阶段和后校正阶段的结构学习表现，以呈现计算量与统计性能之间的权衡。非次高斯和非线性模型的更多评估见附录 E.3-E.4。

为了评估各方法，我们报告平均真阳性率（TPR）、平均错误发现率（FDR）以及以秒为单位的实际运行时间。我们预先计算调优参数集，以确保最大的 λ 对应空图估计，最小的 λ 对应完全连通图估计。我们允许观测值数量 n 和观测变量数量 p 独立变化，因此其中一个可以远大于另一个，以反映高维设置。我们还设置了实际运行时间上限（36 小时）。超过运行时间上限或 FDR 值未定义的情况均标记为缺失值。

6.1 ACV 作为一种可扩展的筛选图估计

我们实现了一个 K 折版本的 ACV，对 K 折进行平均，并将其与相应的 K 折交叉验证估计进行比较。我们考虑 $p=(100、200、500)$，样本量从 $n=50$ 到 $n=5000$，以应对高维设置和计算挑战。我们展示了在高斯数据集下，ACV 用于预筛选和后校正时的结构学习表现，从而分别呈现 ACV 的一致性和将 ACV 与后校正结合后的计算优势。结果见图 1。

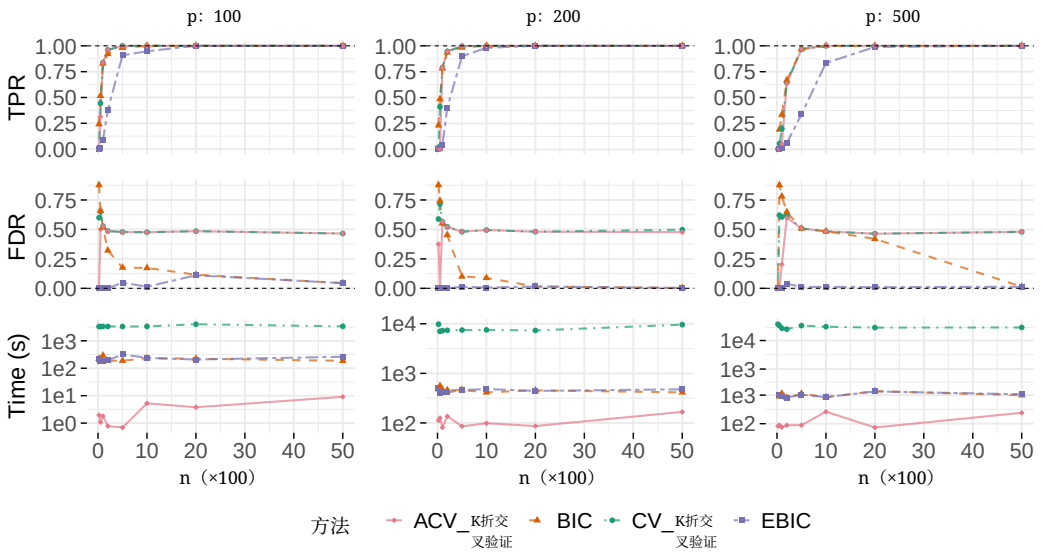


图 1：在预筛选阶段的次高斯数据上学习大规模 K 近邻图时，对 ACV K 折、CV K 折和（扩展）贝叶斯信息准则的评估。

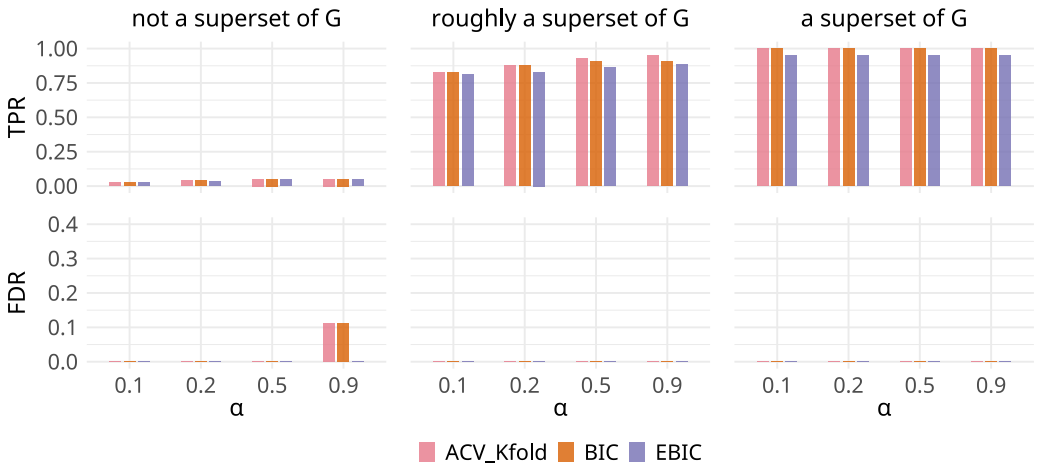


图2：在次高斯数据上的大规模 K 近邻图中，对经 ACV K 折、贝叶斯信息准则和扩展贝叶斯信息准则调优的估计进行后校正后的真正率与错误发现率。

图 1 显示，ACV K 折在真正率方面紧密跟随 K 折交叉验证，同时耗时少得多。这体现了 ACV 在计算方面的作用：它无需反复完整重新拟合，即可获得类似交叉验证的筛选图。同一图还显示了原始交叉验证类方法的统计局限：真正率饱和后，其错误发现率并不会趋于零。这与定理 3.2 和 4.1 一致：ACV 适合作为高召回率筛选图，但不适合作为最终恢复图。贝叶斯信息准则和扩展贝叶斯信息准则提供了重要的似然校准基线；在这些实验中，沿样本量轴增大时，它们往往较晚才纳入真实边，因此有必要对 ACV 筛选结果进行校正，而不是用信息准则取而代之。

6.2 ACV所选图的校正

我们通过将算法 1 应用于图 1 中的筛选图，对校正步骤进行了实证评估。我们使用 $\alpha \in \{0.1, 0.2, 0.5, 0.9\}$，并选择 $n =$（100，500，1000）来代表输入图的三种情形：（1）缺失真实边；（2）近似筛选出真实图；（3）包含真实图。这直接检验了定理 5.2 中的稀疏筛选要求。结果如图 2 所示。

图2 展示了校正的预期效果。在 $n = 100$ 样本量有限时，筛选图会缺失真实边。任何校正步骤都无法恢复这些被遗漏的边，因此真正率（TPR）会降低。在 $n = 500$ 样本量较多时，筛选图近似包含真实图。在这种情况下，校正显著减少了错误发现，同时保留了大多数真实边。在 $n = 1000$ 样本量充足时，筛选图包含真实图，校正会移除额外边，并达到零错误发现率（FDR）。这与理论结果相对应：筛选质量是首要要求，而校正则用于处理原始 CV 型图的过度选择问题。

6.3 弱信号与真实数据

接下来，我们研究边信号接近信号强度条件边界的情形。我们使用高斯数据和 K 近邻图，其中 $k = 1$。根据假设（A3）中的信号强度参数 γ，我们固定 $p = 100$，并将最小偏相关设为随 $n^{-\gamma}$ 变化，其中$\gamma \in \{0, 0.25, 0.5\}$；较大的 γ 对应较弱的信号。

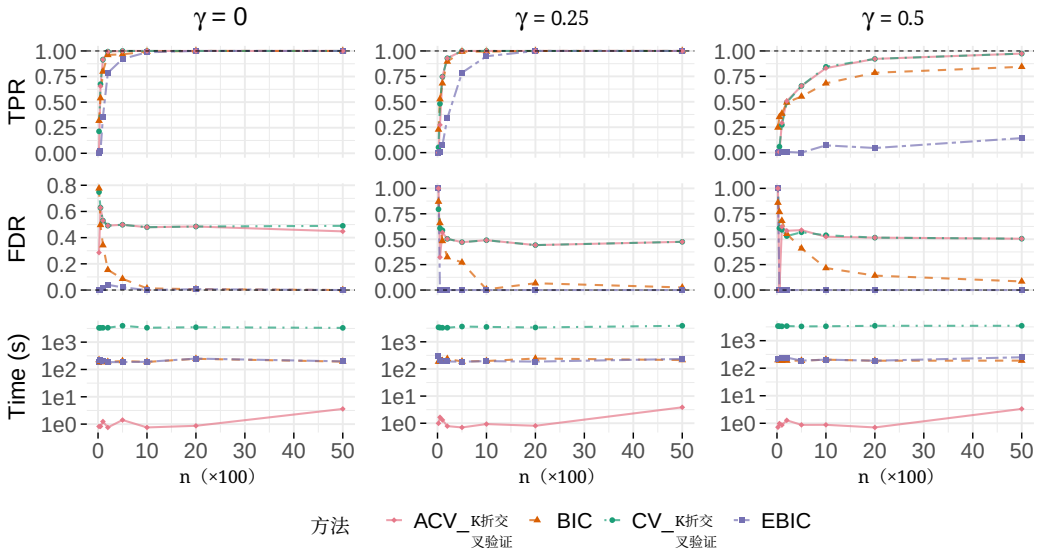


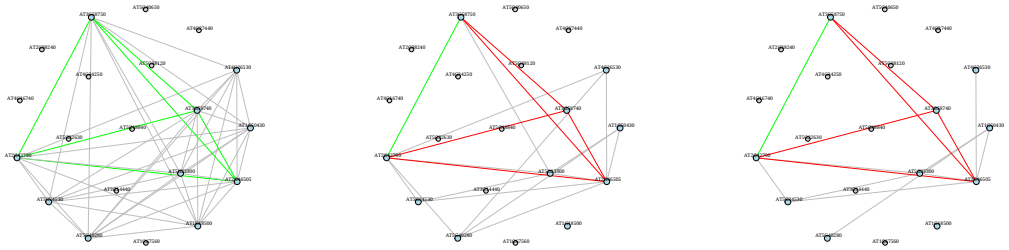
图3：不同信号强度 γ下，ACV K 折、CV K 折与（扩展）BIC 在大型 K 近邻图上的比较。

图3显示，对于强信号或中等强度信号（$\gamma = 0$，0.25），所有方法的真正率都接近完美，错误发现率也非常低，而 ACV 的速度远快于 K 折交叉验证。对于弱信号（$\gamma = 0.5$），扩展贝叶斯信息准则检测到的真实边不到三分之一，贝叶斯信息准则在 $n = 5000$时仍会漏掉真实边。相比之下，K 折交叉验证和 ACV 保留了接近 90% 的真实边，而 ACV 所需时间远少于 K 折交叉验证。这支持了 ACV 作为候选图阶段方法的作用：只有第一阶段的图保留了真实边，校正才能成功。

最后，我们将这些方法应用于一个拟南芥 基因表达数据集，并以已知功能关联构成的图作为参考图 [18, 21, 45]。表1和图4报告了预筛选图任务的结果。ACV及其变体保留了所有已知边，而贝叶斯信息准则和扩展贝叶斯信息准则漏掉了超过80%的已知边。附录 E.3-E.4报告了对数正态数据和非线性伊辛模型下的压力测试；附录 E.6则将这些方法与稳定性选择、TIGER和Lockhart协方差检验进行了比较。

表1：基因表达筛选图比较。

方法	TPR	F1	时间（秒）
BIC	0.167	0.095	0.062
EBIC	0.167	0.125	0.071
ProxACV	**1.000**	**0.240**	56.544
ACV	**1.000**	**0.240**	0.685
ACV K 倍	**1.000**	**0.240**	**0.058**



（a）ACV K 倍　(b) 贝叶斯信息准则　(c) 扩展贝叶斯信息准则

图4：通过 ACV K折、贝叶斯信息准则和扩展贝叶斯信息准则在缩减后的基因表达图上选出的筛选图。 正确识别的边以绿色标示，缺失的边以红色标示。

7 结论与局限

本文研究了高维情形下的图模型选择问题，其中既存在统计方面的问题，也存在计算方面的问题。具体而言，我们考虑使用交叉验证类型的程序。尽管这类程序并不一致，但具有一种颇具吸引力的预筛选性质，可将其用于两阶段筛选与校正策略。这样便可采用近似交叉验证等快速的交叉验证近似方法。我们将ACV固有的不一致性视为一种可扩展的中间估计，它能够编码有用的结构信息。ACV的预筛选性质和不一致性表明，模型选择必须采用后校正程序。我们的理论结果表明，结合后校正步骤后，ACV实际上能够在模型选择中实现一致性，从而解决传统交叉验证方法固有的不一致性和计算问题。这在实践中兼顾了正确性与效率。

我们在算法 1 中提出了这样一种方法，但还有多种后校正策略可用于基于 ACV 构建的中间估计。我们目前的选择依赖于部分独立性检验，并采用更保守的错误发现控制。其他校正方案或许能实现更好的权衡；调整并比较其他方案以寻找最佳权衡，是一个值得探索的未来方向。

总体而言，实证结果验证了所选 ACV 后校正方法的有效性，与现有结构学习方法相比，该方法在一致性和计算效率方面均有显著提升。所提出的方法在大规模图模型中尤其具有优势，因为传统方法往往难以应对其可扩展性问题。我们的工作揭示了 ACV 等不一致程序中隐藏的基本结构信息，并自然地拓展为一种高效且可扩展的结构学习框架，兼顾理论严谨性与实际性能，使其成为高维图模型选择的有力工具。

致谢 BA 获得了美国国家科学基金会 DMS-2610618、IIS-2453378、IIS-1956330 以及 Robert H. Topel 教师研究基金的资助。

参考文献

[1] S. Arlot 和 A. Celisse。用于模型选择的交叉验证程序综述。*StatisticsSurveys*，4:40–79，2010。[2] O. Banerjee、L. El Ghaoui 和 A. d'Aspremont。通过针对多元高斯或二元数据的稀疏最大似然估计进行模型选择。机器学习研究杂志，9:485–516，2008。[3] S. Bates、T. Hastie 和 R. Tibshirani。交叉验证：它估计什么，以及效果如何？《美国统计学会会刊》，119(546):1434–1445，2024。[4] A. Beirami、M. Razaviyayn、S. Shahrampour 和 V. Tarokh。关于参数化学习中的最优泛化能力。载于 I. Guyon、U. V. Luxburg、S. Bengio、H. Wallach、R. Fergus、S. Vishwanathan 和 R. Garnett（编），神经信息处理系统进展，第 30 卷。Curran Associates, Inc.，2017。URL https://proceedings.neurips.cc/paper_files/paper/2017/file/a82d922b133be19c1171534e6594f754-Paper.pdf。[5] P. J. Bickel 和 E. Levina。通过阈值化进行协方差正则化。《统计学年鉴》，36(6):2577–2604，2008。[6] T. T. Cai 和 W. Liu。大规模相关性多重检验。《美国统计学会会刊》，111(513):229–240，2016。doi：10.1080/01621459.2014.999157。[7] G. Csardi 和 T. Nepusz。用于复杂网络研究的 igraph 软件包。*Inter-Journal*，Complex Systems:1695，2006。URL https://igraph.org。[8] A. d'Aspremont、O. Banerjee 和 L. El Ghaoui。稀疏协方差选择的一阶方法。*SIAM* 矩阵分析与应用杂志，30(1):56–66，2008。[9] S. Epskamp。*IsingSampler*：伊辛模型的抽样方法和分布函数，2020。URL https://CRAN.R-project.org/package=IsingSampler。R 软件包版本 0.2.1。[10] J. Fan、H. Liu、Y. Ning 和 H. Zou。用于混合数据的高维半参数潜在图模型。英国皇家统计学会会刊 *B*辑：统计方法论，79 (2):405–421，2017。[11] R. Foygel 和 M. Drton。高斯图模型的扩展贝叶斯信息准则。载于 J. Lafferty、C. Williams、J. Shawe-Taylor、R. Zemel 和 A. Culotta（编），神经信息处理系统进展，第 23 卷。Curran Associates, Inc.，2010。URL https://proceedings.neurips.cc/paper_files/paper/2010/file/072b030ba126b2f4b2374f342be9ed44-Paper.pdf。[12] J. Friedman、T. Hastie 和 R. Tibshirani。使用图形套索进行稀疏逆协方差估计。生物统计学，9(3):432–441，2008。[13] J. Friedman、T. Hastie 和 R. Tibshirani。*glasso*：图形套索：高斯图模型估计，2019。URL https://CRAN.R-project.org/package=glasso。R 软件包版本 1.11。[14] C.-J. Hsieh、M. A. Sustik、I. S. Dhillon 和 P. Ravikumar。QUIC：稀疏逆协方差估计的二次近似。《机器学习研究杂志》，15(83):2911–2947，2014。URL http://jmlr.org/papers/v15/hsieh14a.html。[15] J. Z. Huang、N. Liu、M. Pourahmadi 和 L. Liu。通过惩罚正态似然进行协方差矩阵选择与估计。*Biometrika*，93(1):85–98，2006。[16] B.-Y. Jing、Q.-M. Shao 和 Q. Wang。独立随机变量的自归一化克拉美型大偏差。概率年刊，31(4):2167–2215，2003。doi：10.1214/aop/1068646382。

[17] P. W. W. Koh、K.-S. Ang、H. Teo和P. S. Liang。论影响函数度量群体效应的准确性。载于H. Wallach、H. Larochelle、A. Beygelzimer、F. d'Alché-Buc、E. Fox和R. Garnett编，神经信息处理系统进展，第32卷。Curran Associates, Inc.，2019。URL https://proceedings.neurips.cc/paper_files/paper/2019/file/a78482ce76496fcf49085f2190e675b4-Paper.pdf。[18] I. Lee、B. Ambaru、P. Thakkar、E. M. Marcotte和S. Y. Rhee。利用拟南芥全基因组规模基因网络对基因与性状进行合理关联。《自然-生物技术》，28(2):149–156，2010。[19] J. Lei。带置信度的交叉验证。《美国统计学会会刊》，115 (532):1978–1997，2020。[20] X. Li、T. Zhao、L. Wang、X. Yuan和H. Liu。*flare*：*Lasso*回归家族，2020。URL https://CRAN.R-project.org/package=flare。R 软件包版本1.7.0。[21] Y. Li和S. A. Jackson。通过整合先验生物学知识进行基因网络重建。《*G3*：基因、基因组与遗传学》，5(6):1075–1079，2015。[22] H. Liu和L. Wang。TIGER：一种对调参不敏感的高斯图模型最优估计方法。《电子统计学杂志》，11(1):241–294，2017。[23] H. Liu、J. Lafferty和L. Wasserman。非正态变换：高维无向图的半参数估计。《机器学习研究杂志》，10:2295–2328，2009。[24] H. Liu、K. Roeder和L. Wasserman。高维图模型的正则化选择稳定性方法（stars）。载于J. Lafferty、C. Williams、J. Shawe-Taylor、R. Zemel和A. Culotta编，神经信息处理系统进展，第23卷。Curran Associates, Inc.，2010。URL https://proceedings.neurips.cc/paper_files/paper/2010/file/301ad0e3bd5cb1627a2044908a42fdc2-Paper.pdf。[25] R. Lockhart、J. Taylor、R. J. Tibshirani和R. Tibshirani。Lasso的显著性检验。《统计学年鉴》，42(2):413–468，2014。doi: 10.1214/13-AOS1175。[26] P.-l. Loh和M. J. Wainwright。离散图模型的结构估计：广义协方差矩阵及其逆。载于F. Pereira、C. Burges、L. Bottou和K. Weinberger编，神经信息处理系统进展，第25卷。Curran Associates, Inc.，2012。URL https://proceedings.neurips.cc/paper_files/paper/2012/file/6ba1085b788407963fe0e89c699a7396-Paper.pdf。[27] S. Luo、R. Song和D. Witten。高斯图模型的确定性筛选。*arXiv*预印本 *arXiv:1407.7819*，2014。[28] Z. Lyu、W. M. Tai、M. Kolar和B. Aragam。高斯图模型结构学习中交叉验证的不一致性。载于人工智能与统计国际会议，第3691–3699页。PMLR，2024。[29] T. Ma、Z. Ren和G. C. Tseng。基于多项研究的变量筛选。《中华统计学志》，30 (2):925–953，2020。[30] R. J. Meijer和J. J. Goeman。岭回归的高效近似K 折与留一交叉验证。《生物统计学杂志》，55(2):141–155，2013。[31] N. Meinshausen和P. Bühlmann。基于Lasso的高维图与变量选择。《统计学年鉴》，34(3):1436–1462，2006。[32] Y. N. Moschovakis。描述性集合论。美国数学学会，第2版，2009。URL https://www.math.ucla.edu/~ynm/lectures/dst2009/dst2009.pdf。[33] J. Parraga-Alava、P. Moscato和M. Inostroza-Ponta。*mstknnclust*：*MST-kNN*聚类算法，2023。URL https://CRAN.R-project.org/package=mstknnclust。R 软件包版本0.3.2。

[34] K. R. Rad和A. Maleki。通过近似留一交叉验证对样本外预测误差进行可扩展估计。英国皇家统计学会会刊 *B*辑：统计方法论，82(4):965–996，2020。[35] P. Ravikumar、M. J. Wainwright和J. D. Lafferty。使用 ℓ_1正则化逻辑回归进行高维伊辛模型选择。《统计学年鉴》，38(3):1287–1319，2010。doi: 10.1214/09-AOS691。[36] P. Ravikumar、M. J. Wainwright、G. Raskutti和B. Yu。通过最小化 ℓ_1惩罚对数行列式散度进行高维协方差估计。《电子统计学杂志》，5: 935–980，2011。[37] A. J. Rothman、P. J. Bickel、E. Levina和J. Zhu。稀疏置换不变协方差估计。《电子统计学杂志》，2:494–515，2008。[38] G. Schwarz。估计模型的维数。《统计学年鉴》，6(2):461–464，1978。[39] W. Stephenson和T. Broderick。高维情形下具有理论保证的近似交叉验证。载于人工智能与统计国际会议，第2424–2434页。PMLR，2020。[40] W. Stephenson、M. Udell和T. Broderick。高维低秩数据的近似交叉验证。神经信息处理系统进展，33:9830–9840，2020。[41] M. Stone。交叉验证：综述。统计学：理论与应用统计学杂志，9(1):127–139，1978。[42] M. J. Wainwright。使用ℓ_1约束二次规划（Lasso）进行高维含噪稀疏性恢复的尖锐阈值。*IEEE*信息论汇刊，55 (5):2183–2202，2009。[43] S. Wang、W. Zhou、H. Lu、A. Maleki和V. Mirrokni。用于高维快速参数调优的近似留一法。载于国际机器学习会议，第5228–5237页。PMLR，2018。[44] S. S. Wilks。检验复合假设时似然比的大样本分布。数理统计年鉴，9(1):60–62，1938。doi: 10.1214/aoms/1177732360。[45] A. Wille、P. Zimmermann、E. Vranová、A. Fürholz、O. Laule、S. Bleuler、L. Hennig、A. Preli´c、P. von Rohr、L. Thiele等。拟南芥异戊二烯类基因网络的稀疏图高斯建模。基因组生物学，5(11):R92，2004。[46] A. Wilson、M. Kasy和L. Mackey。近似交叉验证：模型评估与选择的理论保证。载于人工智能与统计国际会议，第4530–4540页。PMLR，2020。[47] W. B. Wu和M. Pourahmadi。纵向数据大型协方差矩阵的非参数估计。*Biometrika*，90(4):831–844，2003。[48] L. Xue和H. Zou。高维非正态图模型的正则化秩估计。《统计学年鉴》，40(5):2541–2571，2012。[49] M. Yuan和Y. Lin。高斯图模型中的模型选择与估计。*Biometrika*，94(1):19–35，2007。[50] T. Zhao、H. Liu、K. Roeder、J. Lafferty和L. Wasserman。用于R中高维无向图估计的huge软件包。《机器学习研究杂志》，13: 1059–1062，2012。URL https://www.jmlr.org。

A ACV估计分析

令 $P \in \mathsf{subG}(\sigma^2)$为第1.1节中定义的零均值分布，并令$\mathbf{X} \sim P^{\otimes n}$ 为$n \times p$ 数据矩阵，其中 $n \geq 2$。记 $\Sigma = \mathsf{Cov}(X) \succ 0$。令$\mathbf{X}^{(i)}$ 为第 i个观测，将其视为列向量；而$\mathbf{X}^{(-i)}$ 表示省略该观测后的矩阵。以下有限样本估计保留了对 p 和 Σ的依赖关系。当 p 和 P随 n变化时，本文的所有渐近结论均采用充分样本量情形，其中明确列出的失败概率趋于零。整个序列中的标准化边缘次高斯界相同。

A.1 $\widetilde{\Theta}^{\lambda,\mathbf{x}}$的推导

对于每个 $i \in [n]$，我们使用全数据估计来近似 $\widehat{\Theta}^{\lambda,\mathbf{x}^{(-i)}}$ 。在第1.1节的总体均值为零假设下，定义

$$F(\Theta, \lambda) := -\log\det\Theta + \mathsf{Tr}(S(\mathbf{X})\Theta) + \lambda\|\Theta\|_1,$$
$$F^{(-i)}(\Theta, \lambda) := -\log\det\Theta + \mathsf{Tr}(S(\mathbf{X}^{(-i)})\Theta) + \lambda\|\Theta\|_1,$$

其中

$$S(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}^{(i)}(\mathbf{X}^{(i)})^\top, \qquad S(\mathbf{X}^{(-i)}) = \frac{1}{n-1}\sum_{j\neq i}\mathbf{X}^{(j)}(\mathbf{X}^{(j)})^\top.$$

因此，

$$\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} = \arg\min_{\Theta\succ 0}\left\{F(\Theta,\lambda) - \mathsf{Tr}\big((S(\mathbf{X}) - S(\mathbf{X}^{(-i)}))\Theta\big)\right\}.$$

令 $\widehat{\mathcal{S}} = \mathsf{supp}(\widehat{\Theta}^{\lambda,\mathbf{x}})$。在 $\widehat{\Theta}^{\lambda,\mathbf{x}}$ 处，光滑似然项的海森矩阵逆为 $\widehat{\Theta}^{\lambda,\mathbf{x}} \otimes \widehat{\Theta}^{\lambda,\mathbf{x}}$。 ^我们在坐标限制之间应用该海森矩阵逆至 $\mathcal{S}$，并定义

$$\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}}_{\widehat{\mathcal{S}}} := \widehat{\Theta}^{\lambda,\mathbf{X}}_{\widehat{\mathcal{S}}} + (\widehat{\Theta}^{\lambda,\mathbf{X}} \otimes \widehat{\Theta}^{\lambda,\mathbf{X}})_{\widehat{\mathcal{S}}\widehat{\mathcal{S}}}(S(\mathbf{X}) - S(\mathbf{X}^{(-i)}))_{\widehat{\mathcal{S}}},$$
$$\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}}_{\widehat{\mathcal{S}}^c} := 0.$$

这里的坐标限制是在向量化的意义下解释的，与正文中的构造相同。这是一次支撑集掩码逆海森矩阵更新。以下分析适用于所展示的构造，并不假设全数据支撑集与留一交叉验证支撑集相等。

A.2 ACV 的评估

我们比较 ACV 与精确留一法交叉验证的验证损失。记 $S_{-i} = S(\mathbf{X}^{(-i)})$、$S = S(\mathbf{X})$ 和 $S_i = S(\mathbf{X}^{(i)})$。定义

$$\mathrm{ACV}(\lambda) := \frac{1}{n}\sum_{i=1}^{n}\left[-\log\det\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} + \mathsf{Tr}(S_i\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}})\right],$$

$$\mathrm{CV}(\lambda) := \frac{1}{n}\sum_{i=1}^{n}\left[-\log\det\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} + \mathsf{Tr}(S_i\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}})\right].$$

T如果任一近似折叠不是正定的，则 ACV 评分为 $+\infty$。在本附录中， ix,

$$\|A\|_1 = \sum_{j,k}|A_{jk}|, \qquad \|A\|_\infty = \max_{j,k}|A_{jk}|$$

表示逐元素范数，而 $\|A\|_F$表示弗罗贝尼乌斯范数。我们用 $\sigma_{\min}(A)$和 $\sigma_{\max}(A)$分别表示最小和最大的奇异值。以下协方差集中结果引自 Ravikumar 等人[36]。

引理 A.1。 对于 $\mathbf{X} \sim P^{\otimes n}$ 以及任意 $i,\ j \in [p]$,

$$\mathsf{Pr}_{\mathbf{X}\sim P^{\otimes n}}\big(|S(\mathbf{X})_{ij} - \Sigma_{ij}| > \varepsilon\big) \leq 4\exp\left\{-\frac{n\varepsilon^2}{128(1+4\sigma^2)^2(\max_j\Sigma_{jj})^2}\right\}$$

对于 $0 < \varepsilon < 8\ (1+4\sigma^2)$ 最大值$_j\,\Sigma_{jj}$。

命题 A.2。 设 $\mathbf{X} \sim P^{\otimes n}$ 并固定 $\lambda > 0$。对于每个 $\delta > 0$,

$$\mathsf{Pr}_{\mathbf{X}\sim P^{\otimes n}}\big(|\operatorname{ACV}(\lambda) - \operatorname{CV}(\lambda)| < \delta\big) \to 1 \qquad \textit{as } n \to \infty.$$

命题 *A.2* 的证明。 我们使用以下逐元素估计。

引理 A.3。 对于任意数据矩阵 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 满足 $n \geq 2$ 的情形、任意 $\lambda > 0$ 以及任意 $i \in [n]$,

$$\begin{aligned}
\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1 &\leq \frac{p^2}{\lambda^2}\|S - S_{-i}\|_\infty, \\
\|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1 &\leq \frac{p^2}{\lambda^2}\|S - S_{-i}\|_\infty, \\
\|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_\infty &\leq \|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_1 \\
&\leq \frac{2p^2}{\lambda^2}\|S - S_{-i}\|_\infty.
\end{aligned}$$

证明见 D.1 节。特别地，该引理不要求全数据估计量与留一法估计量具有相同的支撑集。

令 $C, c > 0$ 表示仅依赖于标准化边缘次高斯参数的常数；它们的取值可能随公式展示而变化。选择 $\varepsilon > 0$ 使其满足

$$\varepsilon \leq c\min\left\{\max_j \Sigma_{jj}, \frac{\lambda}{p}\left(1 + \frac{\lambda}{\max_j \Sigma_{jj}}\right)^{-1/2}\right\}, \qquad \frac{n\varepsilon^2}{(\max_j \Sigma_{jj})^2} \geq 1.$$

根据引理 A.1、边缘次高斯尾界以及并集界，同时事件

$$\max_{j,k\in[p]} |S_{jk} - \Sigma_{jk}| \leq \varepsilon, \qquad\qquad \max_{\substack{i\in[n]\\ j\in[p]}} \big(\mathbf{X}_j^{(i)}\big)^2 \leq \frac{n\varepsilon^2}{\max_j \Sigma_{jj}} \tag{6}$$

的概率至少为

$$1 - Cnp^2 \exp\left\{-\frac{cn\varepsilon^2}{(\max_j \Sigma_{jj})^2}\right\}.$$

在此事件下进行讨论。由于 S 和 S_i 均为半正定，

$$\|S\|_\infty = \max_j S_{jj} \leq 2\max_j \Sigma_{jj}, \qquad \|S_i\|_\infty = \max_j \big(\mathbf{X}_j^{(i)}\big)^2.$$

于是，精确恒等式 $S - S_{-i} = (S_i - S)/(n-1)$ 给出

$$\max_{i\in[n]} \|S - S_{-i}\|_\infty \leq \frac{n\varepsilon^2/(\max_j \Sigma_{jj}) + 2\max_j \Sigma_{jj}}{n-1} \leq \frac{6\varepsilon^2}{\max_j \Sigma_{jj}}.$$

我们首先验证，得分中的所有近似估计量都是正定的。由于惩罚项包含对角元素，对角 KKT 恒等式为

$$\big[(\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1}\big]_{jj} = S_{jj} + \lambda, \qquad \big[(\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}})^{-1}\big]_{jj} = (S_{-i})_{jj} + \lambda.$$

对于任意对称矩阵 B，加权柯西–施瓦茨不等式给出

$$B \succeq -\left(\max_j \big[(\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1}\big]_{jj}\right)\|B\|_1\, \widehat{\Theta}^{\lambda,\mathbf{X}}.$$

将该不等式应用于 $B = \widetilde{\Theta}^{\lambda,\,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\,\mathbf{X}}$，并使用引理 A.3。由此可得

$$\begin{aligned}
\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} &\succeq \left[1 - \frac{p^2(\max_j S_{jj} + \lambda)}{\lambda^2}\|S - S_{-i}\|_\infty\right]\widehat{\Theta}^{\lambda,\mathbf{X}} \\
&\succeq \frac{1}{2}\widehat{\Theta}^{\lambda,\mathbf{X}} \succ 0.
\end{aligned}$$

对于第二个不等式，使用

$$\frac{p^2(\max_j S_{jj}+\lambda)}{\lambda^2}\max_i\|S-S_{-i}\|_\infty \le \frac{Cp^2}{\lambda^2}\left(1+\frac{\lambda}{\max_j \Sigma_{jj}}\right)\varepsilon^2 \le \frac{1}{2},$$

其中，限制条件 ε中的常数 c 取为充分小。

正定性和前述下界意味着

$$\|(\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}})^{-1}\|_\infty \le 2(\max_j S_{jj}+\lambda).$$

事实上，求逆会反转半正定序，而半正定矩阵的最大绝对值元等于其最大的对角元。类似地，$(S_{-i})_{jj} \le nS_{jj}/(n-1) \le 2\max_j S_{jj}$ 得出

$$\|(\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}})^{-1}\|_\infty \le 2\max_j S_{jj}+\lambda \le 2(\max_j S_{jj}+\lambda).$$

令 $B_i = \widetilde{\Theta}^{\lambda,\ \mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\ \mathbf{X}^{(-i)}}$。分别沿两个方向应用对数行列式的凹性可得

$$\begin{aligned}\langle(\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}})^{-1}, B_i\rangle &\le \log\det\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \log\det\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\\ &\le \langle(\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}})^{-1}, B_i\rangle.\end{aligned}$$

因此，

$$\begin{aligned}|\operatorname{ACV}(\lambda)-\operatorname{CV}(\lambda)| &\le \frac{1}{n}\sum_{i=1}^n\left[2(\max_j S_{jj}+\lambda)+\|S_i\|_\infty\right]\|B_i\|_1\\ &\le \frac{2p^2}{\lambda^2 n}\sum_{i=1}^n\left[2(\max_j S_{jj}+\lambda)+\|S_i\|_\infty\right]\|S-S_{-i}\|_\infty. \qquad (7)\end{aligned}$$

这个绝对值界用到了两个对数行列式不等式。

事件 (6) 还意味着

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n\|S_i\|_\infty \le \frac{1}{n}\sum_{i=1}^n \mathsf{Tr}(S_i) = \mathsf{Tr}(S) \le 2p\max_j\Sigma_{jj},\\ &\frac{1}{n}\sum_{i=1}^n\|S_i\|_\infty^2 \le \left(\max_i\|S_i\|_\infty\right)\frac{1}{n}\sum_{i=1}^n\|S_i\|_\infty \le 2np\varepsilon^2. \qquad (8)\end{aligned}$$

将这些估计以及 $\|S-S_{-i}\|_\infty \le (\|S_i\|_\infty + \max_j S_{jj})/(n-1)$ 代入 (7)，我们得到

$$\begin{aligned}|\operatorname{ACV}(\lambda)-\operatorname{CV}(\lambda)| &\le \frac{Cp^3}{\lambda^2}\left[\varepsilon^2+\frac{(\max_j\Sigma_{jj})(\max_j\Sigma_{jj}+\lambda)}{n}\right]\\ &\le \frac{Cp^3}{\lambda^2}\left(1+\frac{\lambda}{\max_j\Sigma_{jj}}\right)\varepsilon^2, \qquad (9)\end{aligned}$$

其中最后一个不等式用到了 $n\varepsilon^2/(\max_j\Sigma_{jj})^2 \ge 1$。

对每个固定的 $\lambda>0$，取

$$\varepsilon = \frac{c}{p^{3/2}n^{1/20}}\min\left\{\max_j\Sigma_{jj}, \frac{\lambda}{\sqrt{1+\lambda/(\max_j\Sigma_{jj})}}\right\}.$$

这同时满足关于 ε 的两个上界限制。每当 $n\varepsilon^2/(\max_j\Sigma_{jj})^2 \ge 1$ 时，前面的界给出

$$|\operatorname{ACV}(\lambda)-\operatorname{CV}(\lambda)| \le Cn^{-1/10}$$

其概率至少为

$$1-Cnp^2\exp\left\{-\frac{cn^{9/10}}{p^3}\min\left\{1, \frac{\lambda^2}{(\max_j\Sigma_{jj})^2+\lambda\max_j\Sigma_{jj}}\right\}\right\}.$$

这展示了多项式维数和协方差因子。在本附录开头所述的充分样本联合极限下，失败概率趋于零；特别地，$n\varepsilon^2/(\max_j\Sigma_{jj})^2 \ge 1$ 最终成立。命题得证。

□

A.3 引理 3.1 的证明

在不致产生歧义时，我们将 $\lambda_*^{\mathbf{X}}$ 和 $\lambda_{\mathrm{ACV}}^{\mathbf{X}}$ 简写为 λ_* 和 λ_{ACV}。对于每个 $\lambda > 0$，定义

$$\begin{aligned} Q(\lambda, \mathbf{X}) &:= \mathsf{E}_{Y\sim P}\Big[-\log\det\widehat{\Theta}^{\lambda,\mathbf{X}} + \mathsf{Tr}(S(Y)\widehat{\Theta}^{\lambda,\mathbf{X}})\Big] \\ &= -\log\det\widehat{\Theta}^{\lambda,\mathbf{X}} + \mathsf{Tr}(\Sigma\widehat{\Theta}^{\lambda,\mathbf{X}}), \\ Q_{\mathbf{Z}}(\lambda, \mathbf{X}) &:= -\log\det\widehat{\Theta}^{\lambda,\mathbf{X}} + \mathsf{Tr}(S(\mathbf{Z})\widehat{\Theta}^{\lambda,\mathbf{X}}). \end{aligned}$$

The 第二个定义适用于维度兼容的任意训练矩阵和评估矩阵 n.

引理 *3.1* 的证明。我们使用在λ_{ACV}的定义之后陈述的完备正选择规则。其存在性将在本证明末尾论证。特别地，对于 ACV 下确界有限的样本，对每个 $\lambda > 0$,

$$\begin{aligned} \mathrm{ACV}(\lambda_{\mathrm{ACV}}) &\le \mathrm{ACV}(\lambda) + n^{-2}, \\ Q(\lambda_*, \mathbf{X}) &\le Q(\lambda, \mathbf{X}) + n^{-2}. \end{aligned}$$

这些不等式包含精确极小值点。

集中性与留一法差值。令 $C, c > 0$ 表示仅依赖于共同标准化边缘次高斯界的常数；其取值在不同公式中可能变化。取

$$0 < \varepsilon' \le \frac{c\,\sigma_{\min}(\Sigma)}{p}, \qquad \frac{n(\varepsilon')^2}{(\max_j \Sigma_{jj})^2} \ge 1,$$

with c 足够小。引理 A.1、边缘次高斯尾界和并集界表明
that

$$\|S - \Sigma\|_\infty \le \varepsilon', \qquad \max_{\substack{i\in[n]\\ j\in[p]}} |\mathbf{X}_{ij}|^2 \le \frac{n(\varepsilon')^2}{\max_j \Sigma_{jj}}$$

以至少为该概率同时成立

$$1 - Cnp^2 \exp\left\{-\frac{cn(\varepsilon')^2}{(\max_j \Sigma_{jj})^2}\right\}.$$

在以下论证中，我们始终在该事件上进行。于是

$$\sigma_{\min}(S) \ge \tfrac{1}{2}\sigma_{\min}(\Sigma), \qquad \mathsf{Tr}(S) \le 2p \max_j \Sigma_{jj}.$$

由于 $S - S_{-i} = (S_i - S)/(n-1)$ 以及 $\|S_i\|_F = \mathsf{Tr}(S_i)$，我们有

$$\begin{aligned} \max_i \|S - S_{-i}\|_F &\le \frac{Cp(\varepsilon')^2}{\max_j \Sigma_{jj}}, \\ \frac{1}{n}\sum_i \|S - S_{-i}\|_F^2 &\le \frac{n^{-1}\sum_i \|S_i\|_F^2}{(n-1)^2} \le \frac{Cp^2(\varepsilon')^2}{n}. \end{aligned}$$

其中，最后一步使用了 $\frac{1}{n}\sum_i \|S_i\|_F^2 \le \max_i \mathsf{Tr}(S_i)\mathsf{Tr}(S) \le 2np^2(\varepsilon')^2$。

与经验风险的全局比较。对于每个固定的 λ，近似更新关于 $S - S_{-i}$ 是线性的，并且对于每个 i 都具有相同的全数据支持集。因此，

$$\frac{1}{n}\sum_i \widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} = \widehat{\Theta}^{\lambda,\mathbf{X}}.$$

此外，$\langle S - S_{-i},\ \widetilde{\Theta}^{\lambda,\,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\,\mathbf{X}}\rangle \ge 0$。对数行列式的凹性以及 $S_i = S + (n-1)(S - S_{-i})$ 因此给出

$$Q_{\mathbf{X}}(\lambda, \mathbf{X}) \le \mathrm{ACV}(\lambda).$$

对于无效候选项，这一点同样成立，其 ACV 值为 $+\infty$。特别是，$S \succ 0$ 意味着 s

$$\mathrm{ACV}(\lambda) \ge Q_{\mathbf{X}}(\lambda, \mathbf{X}) \ge p + \log\det S,$$

因此，在该事件上，ACV 下确界有下界；下面的正比较惩罚项表明它是有限的。

正比较惩罚项。考虑任意

$$0 < \lambda \le \frac{\sigma_{\min}(S)}{np}.$$

该区间仅用于比较，不会限制调参搜索。KKT 条件意味着

$$\|(\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1} - S\|_\infty \le \lambda, \qquad \big[(\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1} - S\big]_{jj} = \lambda.$$

因此

$$\|S^{-1/2}\big((\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1} - S\big)S^{-1/2}\|_F \le \frac{1}{n}.$$

由标量不等式 $(1+t)^{-1} + \log(1+t) - 1 \le 2t^2$ 对于 $|t| \le 1/2$,

$$0 \le Q_{\mathbf{X}}(\lambda, \mathbf{X}) - p - \log\det S \le 2n^{-2}.$$

还可推出

$$\sigma_{\max}(\widehat{\Theta}^{\lambda,\mathbf{X}}) \le \frac{2}{\sigma_{\min}(S)}, \qquad \sigma_{\max}((\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1}) \le 2\sigma_{\max}(S).$$

支撑限制会使弗罗贝尼乌斯范数缩小，因此

$$\|(\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1/2}\big(\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\big)(\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1/2}\|_F \le \frac{8\sigma_{\max}(S)}{\sigma_{\min}(S)^2}\|S - S_{-i}\|_F \le C\left(\frac{p\varepsilon'}{\sigma_{\min}(\Sigma)}\right)^2.$$

这里用到了 $\sigma_{\max}(\Sigma) \le p\max_j \Sigma_{jj}$。将 c 取为足够小的值，可使最终上界至多为$1/2$，因此这些比较惩罚项下的每个近似折叠矩阵都是正定的。

为界定校正，使用 $\sum_i(\widetilde{\Theta}^{\lambda,\,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\,\mathbf{X}}) = 0$ 以及 $0 \le t - \log(1+t) \le t^2$ for $|t| \le 1/2$。因此，ACV 评分与经验风险之间的精确差值满足

$$\begin{aligned} 0 \le\ & \mathrm{ACV}(\lambda) - Q_{\mathbf{X}}(\lambda, \mathbf{X}) \\ \le\ & \left[\frac{4(n-1)}{\sigma_{\min}(S)^2} + \frac{64\sigma_{\max}(S)^2}{\sigma_{\min}(S)^4}\right]\frac{1}{n}\sum_i \|S - S_{-i}\|_F^2 \\ \le\ & C\left(\frac{p\varepsilon'}{\sigma_{\min}(\Sigma)}\right)^2. \end{aligned}$$

第一个系数界定了支撑集限制的克罗内克矩阵的二次型；第二个系数界定了二次对数行列式余项。对于最后一个不等式，ε' 的限制条件意味着

$$n \ge \frac{(\max_j \Sigma_{jj})^2}{(\varepsilon')^2} \ge \frac{p^2(\max_j \Sigma_{jj})^2}{c^2\sigma_{\min}^2(\Sigma)} \ge \frac{\sigma_{\max}^2(\Sigma)}{c^2\sigma_{\min}^2(\Sigma)}.$$

因此，第二个系数至多贡献第一个系数的某个常数倍。

相同的比较惩罚项也满足

$$\|\Sigma^{-1/2}\big((\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1} - \Sigma\big)\Sigma^{-1/2}\|_F \le \frac{p\varepsilon'}{\sigma_{\min}(\Sigma)} + \frac{3}{2n}.$$

应用前述标量似然不等式可得

$$0 \le Q(\lambda, \mathbf{X}) - p - \log\det\Sigma \le C\left[\left(\frac{p\varepsilon'}{\sigma_{\min}(\Sigma)}\right)^2 + n^{-2}\right].$$

在选定惩罚项下的风险比较。在**中选择一个足够大的常数**

$$\varepsilon_n := C\left[\left(\frac{p\varepsilon'}{\sigma_{\min}(\Sigma)}\right)^2 + n^{-2}\right].$$

根据选择不等式和前述比较,

$$\begin{aligned} 0 \le Q_{\mathbf{X}}(\lambda_{\mathrm{ACV}}, \mathbf{X}) - p - \log\det S \le \varepsilon_n, \\ 0 \le Q(\lambda_*, \mathbf{X}) - p - \log\det\Sigma \le \varepsilon_n. \end{aligned}$$

令 c 足够小，可使这些差距至多为 $1/8$；标量函数 $t - \log\ t - 1$ exceeds$1/8$ 在 $[1/2, 2]$之外。将其应用于相应的相对协方差矩阵，可得

$$\sigma_{\max}(\widehat{\Theta}^{\lambda_{\mathrm{ACV}},\mathbf{X}}) \le \frac{4}{\sigma_{\min}(\Sigma)}, \qquad \sigma_{\max}(\widehat{\Theta}^{\lambda_*,\mathbf{X}}) \le \frac{2}{\sigma_{\min}(\Sigma)}.$$

中心化比较由此得出

$$\begin{aligned} -n^{-2} &\le Q(\lambda_{\mathrm{ACV}}, \mathbf{X}) - Q(\lambda_*, \mathbf{X}) \\ &\le \varepsilon_n + \mathsf{Tr}\Big[(\Sigma - S)(\widehat{\Theta}^{\lambda_{\mathrm{ACV}},\mathbf{X}} - \widehat{\Theta}^{\lambda_*,\mathbf{X}})\Big] \\ &\le \varepsilon_n + p\varepsilon' \|\widehat{\Theta}^{\lambda_{\mathrm{ACV}},\mathbf{X}} - \widehat{\Theta}^{\lambda_*,\mathbf{X}}\|_F. \end{aligned}$$

在从 Σ^{-1} 到两个选定估计值的线段上， $-\log\det\Theta + \mathsf{Tr}(\Sigma\Theta)$ 的海森矩阵具有至少为 $\sigma^2_{\min}(\Sigma)/16$ 的弗罗贝尼乌斯曲率，且其梯度在 Σ^{-1} 处为零。根据泰勒定理，

$$\begin{aligned} &\|\widehat{\Theta}^{\lambda_{\mathrm{ACV}},\mathbf{X}} - \widehat{\Theta}^{\lambda_*,\mathbf{X}}\|_F \\ &\quad\le \frac{4\sqrt{2}}{\sigma_{\min}(\Sigma)} \left[\sqrt{Q(\lambda_{\mathrm{ACV}}, \mathbf{X}) - p - \log\det\Sigma} + \sqrt{Q(\lambda_*, \mathbf{X}) - p - \log\det\Sigma}\right]. \end{aligned}$$

代入中心化比较并应用杨氏不等式，可得

$$\begin{aligned} &Q(\lambda_{\mathrm{ACV}}, \mathbf{X}) - p - \log\det\Sigma \\ &\quad\le 3\big[Q(\lambda_*, \mathbf{X}) - p - \log\det\Sigma\big] + 2\varepsilon_n + \frac{64p^2(\varepsilon')^2}{\sigma^2_{\min}(\Sigma)} \\ &\quad\le C\left[\left(\frac{p\varepsilon'}{\sigma_{\min}(\Sigma)}\right)^2 + n^{-2}\right]. \end{aligned}$$

这就证明了所需的总体风险比较。

从风险到惩罚项差值。对于任意 $\Theta \succ 0$ 满足 $-$对数行列式 $\Theta +$迹$(\Sigma\Theta) - p -$对数行列式 $\Sigma \le 1/8$,

$$|(\Theta^{-1} - \Sigma)_{jj}| \le \Sigma_{jj}\sqrt{32\big[-\log\det\Theta + \mathsf{Tr}(\Sigma\Theta) - p - \log\det\Sigma\big]}.$$

两个全数据拟合的对角 KKT 恒等式意味着

$$\lambda_{\mathrm{ACV}} - \lambda_* = \big[(\widehat{\Theta}^{\lambda_{\mathrm{ACV}},\mathbf{X}})^{-1} - (\widehat{\Theta}^{\lambda_*,\mathbf{X}})^{-1}\big]_{jj}.$$

将前述不等式应用于两次拟合可得

$$|\lambda_{\mathrm{ACV}} - \lambda_*| \le C \max_j \Sigma_{jj} \left(\frac{p\varepsilon'}{\sigma_{\min}(\Sigma)} + \frac{1}{n}\right).$$

取 $\varepsilon' = \frac{c\,\sigma_{\min}(\Sigma)}{p(1+\max_j \Sigma_{jj})n^{1/10}}$。这自动满足对 ε' 的上限限制条件。每当$n(\varepsilon')^2/(\max_j \Sigma_{jj})^2 \ge 1$时，它也意味着 $\max_j \Sigma_{jj}/n \le Cn^{-1/5}$，因此

$$|\lambda_{\mathrm{ACV}} - \lambda_*| \le Cn^{-1/10}$$

以至少.....的概率

$$1 - Cnp^2 \exp\left\{-\frac{c\,n^{4/5}\sigma^2_{\min}(\Sigma)}{p^2(1+\max_j \Sigma_{jj})^2(\max_j \Sigma_{jj})^2}\right\}.$$

与先前的论证类似，该结论以趋近于 1 的概率成立。

调参选择的存在性。对于每个 $\lambda > 0$，严格凸性、对数行列式障碍和全元素惩罚项确保拟合结果唯一。预言机目标函数有限，且下界为$p + \log\det\Sigma$，因此其下确界有限。

在 $\lambda = 1 + \max_j S_{jj}$ 处，KKT 条件给出了对角近似折，其中

$$\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}}_{jj} = \frac{\lambda + \frac{n-2}{n-1}S_{jj} + \frac{1}{n-1}(S_i)_{jj}}{(S_{jj}+\lambda)^2} > 0.$$

因此，ACV 存在有限的可行值；当 $S \succ 0$ 其下确界由 $p + \log\det S$ 下界时也是如此。有限下确界都存在正的 n^{-2}-最优选择，而指定的回退方案可处理 ACV 下确界为 $-\infty$ 的情况。

KKT 解映射和折有效性集合都是半代数的。因此，两个得分关于 λ联合博雷尔且分段连续，具有有限多个分段，以及 σ-紧的次水平集和选择截面。博雷尔投影与选择定理保证下确界可测，并给出精确或 n^{-2}-最优选择 [32]。回退方案也具有可测性。

□

B 图恢复评估

本节通过评估 λ_{ACV} 在结构学习中发现真实边的能力，考察其表现。我们将分析分为两部分：首先，在 B.1 节证明定理 4.1，说明 λ_{ACV} 能够渐近地找到所有真实边。接着，在 B.2 节证明定理 3.2，说明仅靠 λ_{ACV} 无法满足一致性。如果不存在歧义，我们将省略对 $\mathbf{X}$ 的依赖。

我们先回顾 KKT 条件 [例如 36]:

引理B.1。 (1)中的凸规划有一个满足 *KKT* 条件的唯一最优解

$$\begin{cases} (\widehat{\Theta}^{\lambda})^{-1}_{ij} - S_{ij} = \lambda \cdot \mathrm{sign}(\widehat{\Theta}^{\lambda}_{ij}) & \widehat{\Theta}^{\lambda}_{ij} \neq 0 \\ |(\widehat{\Theta}^{\lambda})^{-1}_{ij} - S_{ij}| \leq \lambda & \widehat{\Theta}^{\lambda}_{ij} = 0 \end{cases} \tag{10}$$

对任意 $\lambda > 0$。

B.1 定理4.1的证明

证明。 若 $E = \varnothing$，结论显然成立。否则，对于每个 $i \neq j$,

$$\inf_{\substack{\Theta \succ 0 \\ \Theta_{ij}=0}} \{-\log\det\Theta + \mathrm{Tr}(\Sigma\Theta)\} = p + \log\det\Sigma - \log(1-\rho^2_{ij|G}).$$

为验证这个恒等式，只需修改 Σ 的（i, j）和（j, i）两个元素，使相应舒尔补的非对角元为零。其对角元保持不变且为正，因此修改后的矩阵正定。其逆矩阵的（i, j）元素为零，并满足约束一阶方程，因此它是唯一极小值点。修改后的矩阵的行列式为 $\det\Sigma/(1-\rho^2_{ij|G})$，且该极小值点处的迹项等于 p，由此得证。

因此，如果 ACV 选出的拟合结果遗漏了一条真实边 (i, j)，那么由 (A3) 可得

$$\begin{aligned} Q(\lambda_{\mathrm{ACV}}, \mathbf{X}) - p - \log\det\Sigma &\geq -\log(1-\rho^2_{ij|G}) \\ &\geq \rho^2_{ij|G} \geq C_3^2 n^{-2\gamma}. \end{aligned}$$

附录 A.3 中引理 3.1 的证明表明

$$Q(\lambda_{\mathrm{ACV}}, \mathbf{X}) - p - \log\det\Sigma \leq C\left[\left(\frac{p\varepsilon'}{\sigma_{\min}(\Sigma)}\right)^2 + n^{-2}\right]$$

概率至少为

$$1 - Cnp^2\exp\left\{-\frac{cn(\varepsilon')^2}{(\max_j \Sigma_{jj})^2}\right\},$$

给定 $0 < \varepsilon' \le c\,\sigma_{\min}(\Sigma)/p$ 和 $n(\varepsilon')^2/(\max_j \Sigma_{jj})^2 \ge 1$。选择

$$\varepsilon' = \frac{cC_3\sigma_{\min}(\Sigma)}{pn^{\gamma}}$$

当 $c > 0$ 足够小时。由于 $C_3 n^{-\gamma} \le \min_{(i,j)\in E} |\rho_{ij|G}| < 1$，关于 ε' 的上界限制条件成立。另一个限制条件最终会在下文的充分样本量情形下成立。由于 $n^{-2} = o(n^{-2\gamma})$，当 n 足够大时，风险界至多为 $\frac{1}{2}C_3^2 n^{-2\gamma}$。这在同一事件上排除了所有缺失的真实边。因此

$$\mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}\left(G \not\subseteq \widehat{G}^{\lambda_{\mathrm{ACV}},\mathbf{X}}\right) \le Cnp^2 \exp\left\{-\frac{cC_3^2 n^{1-2\gamma}\sigma_{\min}^2(\Sigma)}{p^2(\max_j \Sigma_{jj})^2}\right\} \longrightarrow 0$$

按照全局充分样本量约定。特别地，在 (A2) 下，使最后这个界趋于零的一个充分条件是 $p^2\log(np) = o(n^{1-2\gamma})$。 □

B.2 定理 3.2 的证明

使用附录 A 中的 $S = S(\mathbf{X})$ 和 $S_i = S(\mathbf{X}^{(i)})$。对于 $n \ge p$ 和 $n > 2$，高斯抽样意味着 $S \succ 0$ 以及 S^{-1} 的每个元素几乎必然非零。事实上，S 在正定锥上具有密度，而每个非对角方程 $(S^{-1})_{ij} = 0$ 都定义了一个测度为零的代数集。对于每个这样的样本，KKT 条件给出

$$\|(\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1} - S\|_{\infty} \le \lambda,$$

因此 $\widehat{\Theta}^{\lambda,\,\mathbf{X}} \to S^{-1}$ as $\lambda \downarrow 0$。因此，对于所有充分小的正 λ，其支撑集都是满的。由 (4) 得，

$$\lim_{\lambda\downarrow 0} \widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} = \frac{n-2}{n-1}S^{-1} + \frac{1}{n-1}S^{-1}S_iS^{-1} \succ 0.$$

Thu对于充分小的正 λ，所有近似折叠都是有效的，并且 ACV(λ) 的极限 i s 下确界e.

矩阵行列式引理给出

$$\begin{aligned}
&\lim_{\lambda\downarrow 0} \mathrm{ACV}(\lambda) - p - \log\det S \\
&\quad = -p\log\left(\frac{n-2}{n-1}\right) - \frac{1}{n}\sum_{i=1}^{n}\log\left(1 + \frac{\mathsf{Tr}(S_iS^{-1})}{n-2}\right) \\
&\qquad - \frac{p}{n-1} + \frac{1}{n(n-1)}\sum_{i=1}^{n}\{\mathsf{Tr}(S_iS^{-1})\}^2.
\end{aligned}$$

Here

$$\frac{1}{n}\sum_{i=1}^{n}\mathsf{Tr}(S_iS^{-1}) = p, \qquad \frac{1}{n}\sum_{i=1}^{n}\{\mathsf{Tr}(S_iS^{-1})\}^2 \to p(p+2).$$

第二个断言由 $S^{-1} \to \Sigma^{-1}$ 以及高斯四阶矩的大数定律得到，因为 $\mathsf{E}[(X^\top\Sigma^{-1}X)^2] = p(p+2)$。因此，利用 $0 \le u - \log(1+u) \le u^2/2$ for $u \ge 0$ 可得

$$n\left\{\lim_{\lambda\downarrow 0}\mathrm{ACV}(\lambda) - p - \log\det S\right\} \to p(p+1).$$

该极限仅用于比较正惩罚项下的得分；它并不声称 $\lambda_{\mathrm{ACV}}^{\mathbf{X}}$ 存在任何极限。附录 A.3 给出了对于每个 $\lambda > 0$ 的 $Q_{\mathbf{X}}(\lambda,\ \mathbf{X}) \le \mathrm{ACV}(\lambda)$。由于当 $S \succ 0$ 时，ACV 下确界是有限的，因此所述选择约定同样给出

$$\mathrm{ACV}(\lambda_{\mathrm{ACV}}^{\mathbf{X}}) \le \inf_{\lambda>0}\mathrm{ACV}(\lambda) + n^{-2} \le \lim_{\lambda\downarrow 0}\mathrm{ACV}(\lambda) + n^{-2}.$$

在事件 $\{\widehat{G}^{\lambda^{\mathbf{X}}}{}_{\mathrm{ACV},\mathbf{X}} = G\}$ 上，所选拟合满足 G 的全部零约束。因此，

$$\begin{aligned}
\inf_{\substack{\Theta\succ 0 \\ \Theta_{ij}=0 \text{ if } i<j,\ (i,j)\notin E}} &\{-\log\det\Theta + \mathsf{Tr}(S\Theta)\} \le Q_{\mathbf{X}}(\lambda_{\mathrm{ACV}}^{\mathbf{X}}, \mathbf{X}) \le \mathrm{ACV}(\lambda_{\mathrm{ACV}}^{\mathbf{X}}) \\
&\le \lim_{\lambda\downarrow 0}\mathrm{ACV}(\lambda) + n^{-2}.
\end{aligned}$$

带有这些零约束的高斯模型是维数为 $p+|E|$ 的正则模型，嵌套于维数为 $p(p+1)/2$ 的无约束模型中。真实精度矩阵位于其相对内部，且费舍尔信息在自由坐标上正定。Wilks定理 [44] 给出

$$n\left[\inf_{\substack{\Theta\succ 0\\ \Theta_{ij}=0 \text{ if } i<j,\ (i,j)\notin E}} \{-\log\det\Theta+\mathsf{Tr}(S\Theta)\}-p-\log\det S\right]\xrightarrow{d}\chi^2_{\binom{p}{2}-|E|}.$$

因子 n 是似然比归一化，因为高斯对数似然等于所展示的经验目标函数的 $-n/2$ 倍，二者相差一个加性常数。结合最后三个分量，并利用 χ^2-分布的连续性，可得

$$\limsup_{n\to\infty}\mathbf{Pr}_{\mathbf{X}\sim N(0,\Sigma)^{\otimes n}}\left(\widehat{G}^{\lambda^{\mathbf{X}}_{\mathrm{ACV}},\mathbf{X}}=G\right)\le\mathbf{Pr}\left(\chi^2_{\binom{p}{2}-|E|}\le p(p+1)\right)<1,$$

其中，严格不等式由 $\binom{p}{2}-|E|\ge 1$得出。

C 假发现率控制与一致性

本节证明算法 1 的假阳性控制和选择一致性。附录 A.3 中的全局渐近约定和最终失败概率界意味着

$$p^2\log(np)=O(n^{4/5}),\qquad\frac{p\{\log(np)\}^2}{\sqrt{n}}\longrightarrow 0,\qquad\frac{p^2}{n}\longrightarrow 0.$$

我们在本附录及其辅助结果的证明中始终使用这些推论。若根据$\mathbf{X}\sim P^{\otimes n}$可以明确判断所指对象，我们将简写记号 $\widehat{\Theta}^{\lambda^{\mathbf{X}}}$，$\mathbf{x}\,S(\mathbf{X})$ $\lambda^{\mathbf{X}}_{\mathrm{ACV}}$为 $\widehat{\Theta}^{\lambda}$， S， λ_{ACV}。

C.1 定理 5.2 的证明

Let

$$\mathcal{H}_0:=\{(i,j):1\le i<j\le p,\ (i,j)\notin E\}.$$

^为进行分析，将 $U(i,\ j)$ 和 T_{ij} 在算法 1 中的公式扩展到所有成对组合，包括算法未检验的组合。我们使用以下一致零假设近似，该结果已在 D.2节中证明。

命题 C.1。 设 $\mathbf{X}\sim P^{\otimes n}$，并假设满足假设 *(A1)*、*(A2)* 和 *(A4)*。设 $\widehat{G}$ 是 $\mathbf{X}$的可测^函数，使得 $G\subseteq G$几乎处处成立。于是，对于每个固定的 $M>0$,

$$\sup_{(i,j)\in\mathcal{H}_0}\sup_{0\le t\le M\sqrt{\log(np)}}\left|\frac{\mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(|\widehat{T}_{ij}|\ge t)}{2\{1-\Phi(t)\}}-1\right|\longrightarrow 0.$$

如果 $\mathcal{H}_0$ 为空，则该断言是平凡成立的。

令 $t_\alpha=\Phi^{-1}(1-\alpha/(2p^2))$。对于固定的 $\alpha\in(0,\ 1)$， $t_\alpha=O(\sqrt{\log(np)})$。在定理 5.2 的包含前提下，对确定性集合 $\mathcal{H}_0$ 应用并集界和命题 C.1，可得

$$\begin{aligned}&\mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(\check{G}^{\alpha,\mathbf{X}}\not\subseteq G)\\&\le\sum_{(i,j)\in\mathcal{H}_0}\mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(|\widehat{T}_{ij}|\ge t_\alpha)\\&\le|\mathcal{H}_0|\,\frac{\alpha}{p^2}\{1+o(1)\}\le\alpha\{1+o(1)\}.\end{aligned}$$

取上极限即可证明定理 5.2。

对于一个以趋近于 1 的概率仅包含 G 的输入图，将同样的论证应用于^。^假设输入 $G\cup G$。其输出与 $\{G\subseteq G\}$ 上的实际输出一致，因此

$$\mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(\check{G}^{\alpha,\mathbf{X}}\not\subseteq G)\le\mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(G\not\subseteq\widehat{G})+\alpha\{1+o(1)\}.$$

对于每个固定的 $\varepsilon>0$，同样的论证也适用于 $\alpha=n^{-\varepsilon}$，因为此时 $t_\alpha=O(\sqrt{\log(np)})$.

C.2 定理 5.1 的证明

我们首先为算法 1 使用的拟合残差建立同时界。

引理 C.2。令 $\mathbf{X} \sim P^{\otimes n}$，并假设 *(A1)*、*(A2)* 和 *(A4)*。对于每个 $i \neq j$，令 e_i^0、e_j^0分别表示将 X_i、X_j对 $X_{[p]\setminus\{i,j\}}$进行线性回归所得的总体残差。对于每个条件集

$$\big(\mathrm{ne}(i) \cup \mathrm{ne}(j)\big) \setminus \{i,j\} \subseteq U(i,j) \subseteq [p] \setminus \{i,j\},$$

令 e_i、e_j 表示在 $X_{U(i,j)}$ 上、采用与算法相同的截距约定所得到的拟合的普通最小二乘残差，并记

$$\widehat{r}_{ij} = \frac{1}{n}\sum_{k=1}^{n} e_i^{(k)} e_j^{(k)}, \qquad \widehat{\theta}_{ij} = \frac{1}{n}\sum_{k=1}^{n} (e_i^{(k)} e_j^{(k)} - \widehat{r}_{ij})^2.$$

当 n 足够大时，以至少 $1 - Cp^2/n$ 的概率，对所有此类成对变量和条件集同时有

$$\big|\widehat{r}_{ij} - \mathsf{E}[e_i^0 e_j^0]\big| \le C\sigma_{\min}(\Sigma)\left\{\sqrt{\frac{\log(np)}{n}} + \frac{p\log(np)}{n}\right\},$$

$$c\sigma_{\min}(\Sigma) \le \sqrt{\widehat{\theta}_{ij}} \le C\sigma_{\min}(\Sigma).$$

这里，c、$C > 0$ 仅依赖于共同分布界。此外，

$$\mathsf{E}[e_i^0 e_j^0] = \rho_{ij|G}\sqrt{\mathsf{Var}(e_i^0)\,\mathsf{Var}(e_j^0)}, \qquad \sigma_{\min}(\Sigma) \le \mathsf{Var}(e_i^0), \mathsf{Var}(e_j^0) \le \sigma_{\max}(\Sigma).$$

该引理的证明见 D.3 节。

令 $\alpha = n^{-\varepsilon}$，其中 $\varepsilon > 2\tau$ 固定，并将 $\widehat{G} = \widehat{G}^{\lambda_{\mathrm{ACV}},}$ $\mathbf{x}$ 作为算法 1 的输入。将其输出记为 $\check{G}^{\alpha,}$ $\mathbf{x}$，如附录 C.1 所示。附录 C.1 中证明的界与定理 4.1 给出

$$\mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(\check{G}^{\alpha,\mathbf{X}} \not\subseteq G) \le \mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(G \not\subseteq \widehat{G}) + \alpha\{1 + o(1)\} \longrightarrow 0.$$

若 $E = \varnothing$，则定理得证。否则，还需证明每条真实边都被保留。

$\widehat{\ }$在 $\{G \subseteq G\}$ 上，每条真实边都会被检验，且其条件集满足前述引理。由 (A3) 及引理中的总体恒等式，

$$\min_{(i,j)\in E} |\mathsf{E}[e_i^0 e_j^0]| \ge C_3\sigma_{\min}(\Sigma)n^{-\gamma}.$$

由全局渐近约定和附录 A.3 可得 $p^2\log(np) = O(n^{4/5})$。结合 (A1) 和 $0 \le \gamma < 1/2$，可推出

$$n^{\gamma}\left\{\sqrt{\frac{\log(np)}{n}} + \frac{p\log(np)}{n}\right\} \le C\left(n^{\gamma-1/2} + n^{\gamma-3/5}\right)\sqrt{\log(np)} \longrightarrow 0.$$

因此，在引理所述的同时事件上，对所有充分大的 n，

$$\min_{(i,j)\in E} |\widehat{r}_{ij}| \ge \frac{C_3}{2}\sigma_{\min}(\Sigma)n^{-\gamma}, \qquad \min_{(i,j)\in E} |\widehat{T}_{ij}| \ge cn^{1/2-\gamma}.$$

第二个不等式用到了引理中 $\sqrt{\widehat{\theta}_{ij}}$ 的正下界和上界。检验阈值满足

$$t_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2p^2}\right) \le \sqrt{2\log(2p^2/\alpha)} = O(\sqrt{\log n}) = o(n^{1/2-\gamma}).$$

因此，在引理中的筛选事件与同时事件的交集上，每条真实边都会被保留。因此

$$\mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(G \not\subseteq \check{G}^{\alpha,\mathbf{X}}) \le \mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(G \not\subseteq \widehat{G}) + \frac{Cp^2}{n} \longrightarrow 0,$$

其中 $p^2/n \to 0$ 可由同一附录 A.3 中的界推出。结合这两个错误概率即可证明该定理。

D 技术引理

D.1 引理 A.3 的证明

引理 D.1。 对于每个数据矩阵 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 满足 $n \geq 2$ 的情形，对于任意 $\lambda > 0$ 和任意 $i \in [n]$,

$$\begin{aligned}
\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1 &\leq \frac{p^2}{\lambda^2}\|S - S_{-i}\|_\infty, \\
\|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1 &\leq \frac{p^2}{\lambda^2}\|S - S_{-i}\|_\infty, \\
\|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_\infty &\leq \|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_1 \\
&\leq \frac{2p^2}{\lambda^2}\|S - S_{-i}\|_\infty.
\end{aligned}$$

证明。 回想一下， $\|\cdot\|_1$ 表示逐元素 ℓ_1 范数。根据 KKT 条件,

$$\begin{aligned}
\mathsf{Tr}\left(S\widehat{\Theta}^{\lambda,\mathbf{X}}\right) + \lambda\|\widehat{\Theta}^{\lambda,\mathbf{X}}\|_1 = p, \\
\mathsf{Tr}\left(S_{-i}\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\right) + \lambda\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_1 = p.
\end{aligned}$$

由于 $S, S_{-i} \succeq 0$ 且两个估计量都是正定的,

$$\|\widehat{\Theta}^{\lambda,\mathbf{X}}\|_1 \leq \frac{p}{\lambda}, \qquad \|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_1 \leq \frac{p}{\lambda}.$$

根据加权柯西–施瓦茨不等式,

$$\begin{aligned}
&\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1^2 \\
&\quad\leq \|\widehat{\Theta}^{\lambda,\mathbf{X}}\|_1\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_1 \times \mathsf{Tr}\left[\left((\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1} - (\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}})^{-1}\right)\left(\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\right)\right].
\end{aligned}$$

此外，KKT 条件和 ℓ_1 次梯度的单调性意味着

$$\begin{aligned}
&\mathsf{Tr}\left[\left((\widehat{\Theta}^{\lambda,\mathbf{X}})^{-1} - (\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}})^{-1}\right)\left(\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\right)\right] \\
&\quad\leq \mathsf{Tr}\left[(S - S_{-i})\left(\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\right)\right] \\
&\quad\leq \|S - S_{-i}\|_\infty\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1.
\end{aligned}$$

结合前面的不等式，可得

$$\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1^2 \leq \frac{p^2}{\lambda^2}\|S - S_{-i}\|_\infty\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1.$$

当最后一个范数非零时约去它；范数为零的情形显然成立，因此得到

$$\begin{aligned}
\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_\infty &\leq \|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1 \\
&\leq \frac{p^2}{\lambda^2}\|S - S_{-i}\|_\infty. \qquad (11)
\end{aligned}$$

^对于近似更新，将 $\mathcal{S}$中所有坐标的绝对值求和。根据其定义 ,

$$\begin{aligned}
&\|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1 \\
&\quad= \|\left(\widehat{\Theta}^{\lambda,\mathbf{X}} \otimes \widehat{\Theta}^{\lambda,\mathbf{X}}\right)_{\widehat{\mathcal{S}}\widehat{\mathcal{S}}}(S - S_{-i})_{\widehat{\mathcal{S}}}\|_1 \\
&\quad\leq \left[\sum_{j\in\widehat{\mathcal{S}}}\sum_{k\in\widehat{\mathcal{S}}}|\left(\left(\widehat{\Theta}^{\lambda,\mathbf{X}} \otimes \widehat{\Theta}^{\lambda,\mathbf{X}}\right)_{\widehat{\mathcal{S}}\widehat{\mathcal{S}}}\right)_{jk}|\right]\|S - S_{-i}\|_\infty \\
&\quad\leq \|\widehat{\Theta}^{\lambda,\mathbf{X}} \otimes \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1\|S - S_{-i}\|_\infty \\
&\quad= \|\widehat{\Theta}^{\lambda,\mathbf{X}}\|_1^2\|S - S_{-i}\|_\infty \leq \frac{p^2}{\lambda^2}\|S - S_{-i}\|_\infty.
\end{aligned}$$

结合前面对 $\|\widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1$ 的界，三角不等式给出

$$\begin{aligned}\|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_\infty &\le \|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_1 \\ &\le \|\widetilde{\Theta}^{\lambda,\mathbf{X}^{(-i)}} - \widehat{\Theta}^{\lambda,\mathbf{X}}\|_1 + \|\widehat{\Theta}^{\lambda,\mathbf{X}} - \widehat{\Theta}^{\lambda,\mathbf{X}^{(-i)}}\|_1 \\ &\le \frac{2p^2}{\lambda^2}\|S - S_{-i}\|_\infty.\end{aligned}$$

这就证明了该引理。

□

D.2 命题 C.1 的证明

命题 D.2。 令 $\mathbf{X} \sim P^{\otimes n}$ 并假设假设 *(A1)*、*(A2)* 和 *(A4)*。令 $\widehat{G}$ 为 $\mathbf{X}$的一个可测函数，使得 $G \subseteq G$几乎处处成立。则对于每个固定的 $M > 0$,

$$\sup_{(i,j)\in\mathcal{H}_0} \sup_{0\le t\le M\sqrt{\log(np)}} \left| \frac{\mathbf{Pr}_{\mathbf{X}\sim P^{\otimes n}}(|\widehat{T}_{ij}| \ge t)}{2\{1-\Phi(t)\}} - 1 \right| \longrightarrow 0.$$

若 $\mathcal{H}_0$ 为空，则断言是空泛成立的。

证明。 固定 $M > 0$并假设 $\mathcal{H}_0$ 非空。全局渐近约定和附录 A.3 中的最终界蕴含 $p^2\log(np) = O(n^{4/5})$，因此

$$\frac{p\{\log(np)\}^2}{\sqrt{n}} \longrightarrow 0.$$

对于每个固定的条件集，共同缩放不会改变检验统计量。由于下述界对这些集合一致成立，不失一般性地，我们处理 $\sigma_{\min}(\Sigma) = 1$ 和 $\sigma_{\max}(\Sigma) \le C_2$。
令 ε_i 为假设（A4）中的总体误差。局部马尔可夫性质共同给出$\varepsilon_i \perp X_{[p]\setminus\{i\}}$。对于一个零假设对， ε_j 是 $X_{[p]\setminus\{i\}}$的函数，因此 $\varepsilon_i \perp \varepsilon_j$。因此，

$$\theta_{ij} := \mathsf{Var}(\varepsilon_i\varepsilon_j) = \mathsf{Var}(\varepsilon_i)\mathsf{Var}(\varepsilon_j), \qquad 1 \le \theta_{ij} \le C_2^2.$$

这里， $1 \le \mathsf{Var}(\varepsilon_i) \le \Sigma_{ii} \le C_2$ 由协方差界和回归正交性推出。独立性和中心化也意味着 $\mathrm{E}e^{t\varepsilon_i} \le \mathrm{E}e^{tX_i}$，因此总体误差的次高斯参数一致有界。

我们首先控制由这些误差构成的经验量。以下所有样本和均从 1 求和至 n，通用常数 C_M 仅取决于 M 和共同的分布界。令 A_1 为满足以下条件的事件：

$$\begin{aligned}\max_{i,j}|S_{ij} - \Sigma_{ij}| &\le C_M\sqrt{\frac{\log(np)}{n}}, \\ \max_i\left(\left|\frac{1}{n}\sum_k X_i^{(k)}\right| + \left|\frac{1}{n}\sum_k \varepsilon_i^{(k)}\right|\right) &\le C_M\sqrt{\frac{\log(np)}{n}}, \\ \max_{i,\,\ell\ne i}\left|\frac{1}{n}\sum_k X_\ell^{(k)}\varepsilon_i^{(k)}\right| &\le C_M\sqrt{\frac{\log(np)}{n}}, \\ \max_{i,k}(|X_i^{(k)}| + |\varepsilon_i^{(k)}|) &\le C_M\sqrt{\log(np)},\end{aligned}$$

并且，对每个 $(i,\ j) \in \mathcal{H}_0$,

$$\frac{\left|n^{-1}\sum_k \varepsilon_i^{(k)}\varepsilon_j^{(k)}\right|}{\sqrt{\theta_{ij}}} + \left|\frac{\sum_k(\varepsilon_i^{(k)}\varepsilon_j^{(k)})^2}{n\theta_{ij}} - 1\right| \le C_M\sqrt{\frac{\log(np)}{n}}.$$

用 A_1^c 表示其补集。次高斯尾界和伯恩斯坦不等式可控制协方差项、样本均值、得分、最大值以及总体乘积均值。对于总体平方

乘积，尾界 $\mathbf{Pr}((\varepsilon_i\varepsilon_j)^2 > x) \le C\exp(-c\sqrt{x})$ 允许在 $C_M\{\log(np)\}^2$ 处进行截断。随后，伯恩斯坦不等式给出的偏差量级为 $\sqrt{\log(np)/n} + \{\log(np)\}^3/n$。在（A1）条件下，不等式的第二项可由第一项吸收，而截断概率和偏差均可任意减小。选择常数，使每个失败概率都不超过 $(np)^{-(M+1)^2-6}$，并对至多 Cnp^2 项应用并集界，可得

$$\mathbf{Pr}(A_1^c) \le C_M (np)^{-(M+1)^2-4}.$$

在 A_1 上，当 n 足够大时，完整的经验格拉姆矩阵的最小特征值至少为 $1 - C_M p\sqrt{\log(np)}/n \ge 1/2$。样本均值界在添加截距列后给出相同结论。每个回归格拉姆矩阵都是该完整矩阵的主子矩阵，因此逆矩阵界对每个条件集同时成立。对于任意零假设对以及任意 $\mathrm{ne}(i) \cup \mathrm{ne}(j) \subseteq U \subseteq [p] \setminus \{i, j\}$，总体预测变量都属于回归空间。因此，$\varepsilon_i - e_i$ 是样本总体误差向量 ε_i 的普通最小二乘投影。正规方程给出

$$\frac{\|\varepsilon_i - e_i\|_2^2}{n} \le 2\left\{\left|\frac{1}{n}\sum_k \varepsilon_i^{(k)}\right|^2 + \sum_{\ell \in U(i,j)}\left|\frac{1}{n}\sum_k X_\ell^{(k)}\varepsilon_i^{(k)}\right|^2\right\} \le C_M\frac{p\log(np)}{n},$$

$$\max_k |\varepsilon_i^{(k)} - e_i^{(k)}| \le C_M\sqrt{p\log(np)}\sqrt{\frac{p\log(np)}{n}} = C_M\frac{p\log(np)}{\sqrt{n}}.$$

样本均值项涵盖截距；否则可以省略。第二行还使用了每个设计行的界。我们注意到，对于任意 $j \in [p]$，相同的估计仍然成立。

令 $\widehat{r}_{ij} = n^{-1}\sum_k e_i^{(k)}e_j^{(k)}$ 和 $\widehat{\theta}_{ij} = n^{-1}\sum_k (e_i^{(k)}e_j^{(k)} - \widehat{r}_{ij})^2$。与公共回归空间正交可得

$$\widehat{r}_{ij} - \frac{1}{n}\sum_k \varepsilon_i^{(k)}\varepsilon_j^{(k)} = -\frac{1}{n}(\varepsilon_i - e_i)^\top(\varepsilon_j - e_j),$$

其绝对值至多为 $C_M p\log(np)/n$。展开乘积差并利用投影界，还可得到

$$\left\{\frac{1}{n}\sum_k (e_i^{(k)}e_j^{(k)} - \varepsilon_i^{(k)}\varepsilon_j^{(k)})^2\right\}^{1/2} \le C_M\left\{\frac{\sqrt{p}\log(np)}{\sqrt{n}} + \frac{p^{3/2}\{\log(np)\}^{3/2}}{n}\right\} \le C_M\frac{\sqrt{p}\log(np)}{\sqrt{n}}.$$

这就界定了两个经验标准差之间的差异。此外，总体方差减去经验方差为 $\theta_{ij}\{1 + O(\sqrt{\log(np)/n})\}$，在 A_1 上成立。因此，

$$\left|\frac{\widehat{\theta}_{ij}}{\theta_{ij}} - 1\right| \le C_M\frac{\sqrt{p}\log(np)}{\sqrt{n}}.$$

对于每个符合条件的 $U(i, j)$，这些比较都在同一事件上成立。

将 Z_{ij} 中的 e_i, e_j 替换为 $\varepsilon_i, \varepsilon_j$，定义 Z_{ij}。只要其分母为正，

$$|\widehat{Z}_{ij}| \ge t \quad \Longleftrightarrow \quad \frac{\left|\sum_k \varepsilon_i^{(k)}\varepsilon_j^{(k)}\right|}{\left\{\sum_k (\varepsilon_i^{(k)}\varepsilon_j^{(k)})^2\right\}^{1/2}} \ge \frac{t}{\sqrt{1 + t^2/n}}.$$

对于每个零假设对，标准化总体乘积在各个样本间独立同分布，均值为零、方差为一，且三阶绝对矩一致有界。Jing等人 [16] 的自归一化中等偏差定理给出该比值的正态尾部近似，其相对误差为 $O((1+t)^3/\sqrt{n})$。阈值变换使正态尾部发生相对 $O((1+t)^4/n)$的变化。对于 $0 \le t \le (M+1)\sqrt{\log(np)}$，这两种误差均一致趋于零。分母为零的事件包含于 A_1^c forlarge n，并且

$$\frac{\mathbf{Pr}(A_1^c)}{1 - \Phi((M+1)\sqrt{\log(np)})} \le C_M\sqrt{\log(np)}\,(np)^{-(M+1)^2/2-4} \longrightarrow 0.$$

由此可知，对于零假设对以及该范围内的 t，一致地有

$$\mathbf{Pr}(|\widehat{Z}_{ij}| \geq t) = 2\{1-\Phi(t)\}\{1+o(1)\}.$$

最后，在 A_1上，两种经验方差都远离零。上述分子与方差的比较意味着

$$|\widehat{T}_{ij} - \widehat{Z}_{ij}| \leq \delta_n(1+|\widehat{Z}_{ij}|), \qquad \delta_n = C_M \frac{p\log(np)}{\sqrt{n}}.$$

S $\widehat{\ }$由于 $G \subseteq G$，这适用于算法所用的条件集。对于足够大的 n，

$$\begin{aligned}
&\mathbf{Pr}\left(|\widehat{Z}_{ij}| \geq \frac{t+\delta_n}{1-\delta_n}\right) - \mathbf{Pr}(A_1^c) \\
&\leq \mathbf{Pr}(|\widehat{T}_{ij}| \geq t) \\
&\leq \mathbf{Pr}\left(|\widehat{Z}_{ij}| \geq \frac{\max\{t-\delta_n, 0\}}{1+\delta_n}\right) + \mathbf{Pr}(A_1^c).
\end{aligned}$$

正态分布上尾的对数导数为 $O(1+t)$，且 $\delta_n\{1+M\sqrt{\log(np)}\}^2 \to 0$。因此，对于 $0 \leq t \leq M\sqrt{\log(np)}$，两个平移正态分布尾部均一致等价于 $2\{1-\Phi(t)\}$。根据前述界，异常概率可忽略。将夹逼关系除以 $2\{1-\Phi(t)\}$，即可证明该命题。 □

D.3 引理 C.2 的证明

引理 D.3。设 $\mathbf{X} \sim P^{\otimes n}$，并假设 *(A1)*、*(A2)* 和 *(A4)*。对于每个 $i \neq j$，令 e_i^0, e_j^0 为将 X_i，X_j 对 $X_{[p]\setminus\{i,j\}}$进行线性回归所得的总体残差。对于每个条件集

$$\big(\mathrm{ne}(i) \cup \mathrm{ne}(j)\big) \setminus \{i,j\} \subseteq U(i,j) \subseteq [p] \setminus \{i,j\},$$

令 e_i, e_j 为在 $X_{U(i,j)}$ 上拟合的普通最小二乘残差，采用与算法相同的截距约定，并记

$$\widehat{r}_{ij} = \frac{1}{n}\sum_{k=1}^{n} e_i^{(k)} e_j^{(k)}, \qquad \widehat{\theta}_{ij} = \frac{1}{n}\sum_{k=1}^{n} (e_i^{(k)} e_j^{(k)} - \widehat{r}_{ij})^2.$$

当 n 充分大时，以至少 $1 - Cp^2/n$ 的概率，对所有此类配对和条件集同时成立，

$$\left|\widehat{r}_{ij} - \mathsf{E}[e_i^0 e_j^0]\right| \leq C\sigma_{\min}(\Sigma)\left\{\sqrt{\frac{\log(np)}{n}} + \frac{p\log(np)}{n}\right\},$$

$$c\sigma_{\min}(\Sigma) \leq \sqrt{\widehat{\theta}_{ij}} \leq C\sigma_{\min}(\Sigma).$$

此处 $c, C > 0$ 仅取决于共同分布界。此外，

$$\mathsf{E}[e_i^0 e_j^0] = \rho_{ij|G}\sqrt{\mathsf{Var}(e_i^0)\,\mathsf{Var}(e_j^0)}, \qquad \sigma_{\min}(\Sigma) \leq \mathsf{Var}(e_i^0), \mathsf{Var}(e_j^0) \leq \sigma_{\max}(\Sigma).$$

证明。这些界在共同缩放下具有齐次性，因此只需在$\sigma_{\min}(\Sigma) = 1$ 且 $\sigma_{\max}(\Sigma) \leq C_2$的条件下证明。与 D.2节 中一样，全局渐近约定意味着

$$p^2\log(np) = O(n^{4/5}), \qquad \frac{p\{\log(np)\}^2}{\sqrt{n}} \longrightarrow 0.$$

在整个证明过程中，$c, C > 0$ 仅取决于共同分布界。

令 ε_i 为假设（A4）中的总体误差。局部马尔可夫性质与该假设可得 $\varepsilon_i \perp X_{[p]\setminus\{i\}}$，这一点已在 D.2节中证明。完整的逐节点回归系数为$_^{(\Sigma^{-1})_{i,\ -i}\,(\Sigma^{-1})_{ii}}$，并且在真实邻域之外均为零。分块正规方程给出

$$\begin{pmatrix} e_i^0 \\ e_j^0 \end{pmatrix} = \big[(\Sigma^{-1})_{\{i,j\},\{i,j\}}\big]^{-1}\begin{pmatrix} (\Sigma^{-1})_{ii}\varepsilon_i \\ (\Sigma^{-1})_{jj}\varepsilon_j \end{pmatrix},$$

$$\mathsf{Cov}\begin{pmatrix} e_i^0 \\ e_j^0 \end{pmatrix} = \big[(\Sigma^{-1})_{\{i,j\},\{i,j\}}\big]^{-1}.$$

相应的回归系数在 (ne(i) $\cup$ ne(j)) $\setminus \{i,j\}$ 之外均为零。因此，对于引理中的每个条件集，这些也都是总体残差。协方差恒等式给出了所述的总体相关系数和方差界。

乘在（ε_i，ε_j）上的系数矩阵的算子范数至多为 $\kappa(\Sigma) \le C_2$。因此，D.2节中为节点误差建立的次高斯界也一致适用于 e_i^0，e_j^0。它们的样本均值、协变量得分和最大值，都是相应节点误差量的有界线性组合。因此，D.2节中使用的协方差、样本均值、得分和最大值界表明，除概率至多为 $C(np)^{-2}$ 的事件外，

$$\frac{\|e_i^0 - e_i\|_2}{\sqrt{n}} \le C\sqrt{\frac{p\log(np)}{n}},$$
$$\max_k |(e_i^0)^{(k)} - e_i^{(k)}| \le C\frac{p\log(np)}{\sqrt{n}},$$
$$\max_k |(e_i^0)^{(k)}| \le C\sqrt{\log(np)},$$

并且对 j 也成立相同的界，且这些界同时适用于所有成对组合及符合条件的条件集。向量表达式中的总体残差表示其样本向量。为完整起见，前两个界可由正规方程推出：得分向量的欧几里得范数至多为$C\sqrt{p}\log(np)/n$，Gram矩阵的逆的算子范数至多为 2，且每个设计行的欧几里得范数至多为 $C\sqrt{p}\log(np)$。完整的 Gram 界适用于所有主子矩阵；纳入常数列时，样本均值界也给出相同结果。因此，该论证适用于任一种截距约定。

与公共回归空间正交可得

$$\widehat{r}_{ij} - \frac{1}{n}\sum_{k=1}^{n}(e_i^0)^{(k)}(e_j^0)^{(k)} = -\frac{1}{n}(e_i^0 - e_i)^\top(e_j^0 - e_j).$$

其绝对值至多为 $Cp\log(np)/n$。总体成对乘积是次指数的，因此由伯恩斯坦不等式以及对所有配对取并集界可得

$$\max_{i\neq j}\left|\frac{1}{n}\sum_{k=1}^{n}(e_i^0)^{(k)}(e_j^0)^{(k)} - \mathsf{E}[e_i^0 e_j^0]\right| \le C\sqrt{\frac{\log(np)}{n}}$$

除去概率至多为 $C(np)^{-2}$ 的事件。由此证明了 $\widehat{r}_{ij}$ 的所述界。

为控制分母，节点回归方程给出

$$e_i^0 = -\frac{(\Sigma^{-1})_{ij}}{(\Sigma^{-1})_{ii}}e_j^0 + \varepsilon_i, \qquad \varepsilon_i \perp e_j^0,$$

因为 e_j^0 是 $X_{[p]\setminus\{i\}}$的函数。由于 ε_i 是中心化的，

$$\begin{aligned}\mathsf{Var}(e_i^0 e_j^0) &= \frac{(\Sigma^{-1})_{ij}^2}{(\Sigma^{-1})_{ii}^2}\mathsf{Var}((e_j^0)^2) + \mathsf{Var}(\varepsilon_i)\mathsf{Var}(e_j^0)\\ &\ge 1.\end{aligned}$$

次高斯矩界还给出 Var($e_i^0 e_j^0$) $\le C$ and E$|e_i^0 e_j^0|^4 \le C$。因此，对经验一阶和二阶乘积矩应用切比雪夫不等式，再对所有配对取并集界，可知总体乘积的所有中心化经验方差都介于两个正的常数之间，除去概率至多为 Cp^2/n 的事件。

展开拟合乘积之差，并利用前述投影界和最大值界，可得

$$\left\{\frac{1}{n}\sum_{k=1}^{n}\left(e_i^{(k)}e_j^{(k)} - (e_i^0)^{(k)}(e_j^0)^{(k)}\right)^2\right\}^{1/2} \le C\left\{\frac{\sqrt{p}\log(np)}{\sqrt{n}} + \frac{p^{3/2}\{\log(np)\}^{3/2}}{n}\right\} \longrightarrow 0,$$

该结论对所有符合条件的条件集均一致成立。中心化会使欧几里得范数收缩，因此这也界定了拟合乘积与总体乘积的经验标准差之差。

因此 $c \le \sqrt{}$

$\widehat{\theta}_{ij} \le C$ 同时在同一事件上成立，且 n 足够大。合并各个例外事件的概率，并利用 ($np)^{-2} \le p^2/n$ 可知，总失败概率至多为 Cp^2/n。重新缩放即可完成该引理。 □

E 实验细节

E.1 设置

计算 所有实验均在一台 16 核英特尔至强金牌 6248R 处理器上进行，其主频为 3.00 GHz，内存为 16 GB。所有实验的运行时间上限均为 36 小时。

图模型我们使用了四种常见的图集：带状图、无标度图（SF）、埃尔德什-雷尼图（ER）和 k近邻图（KNN）。带状图和无标度图（SF）的图结构遵循 flare [20]包，而 ER 图和 KNN 图则分别在 R 中使用 i-graph [7] 和 mstknnclust [33] 生成。具体而言，图初始化如下：

- 带状图：给定观测值数量 n 和变量数量 p，这些图是单链图； u， v 在 r-flare 包中设为默认值，带宽设为 1。
- 无标度图（SF）：这些图由巴拉巴西–阿尔伯特模型生成；图以两个相连的节点初始化，新节点连接到某个现有节点的概率与该现有节点的度成正比。
- 埃尔德什-雷尼图（ER）：这些是具有 p 个节点和 $\lceil {}^{p}{}_{\log_{10}{}^{p}} \rceil$ 条边的随机无向图，这些边从所有可能边的集合中均匀随机选取。
- k近邻图（KNN）：在某些随机图中，只有当节点是彼此距离对应的 k近邻之一时，它们才会相连。生成一个均匀分布的 $p \times p$ 矩阵作为随机数据，用于计算节点之间的欧几里得距离。我们设置 $k = 1$ 来构建稀疏图。

方法 我们考虑以下方法来调整 λ 。

- 留一交叉验证（LOOCV）：每次移除一个数据点作为验证集，其余数据点用作训练集。
- K折交叉验证（CV K折）：我们设置 $K = 3$，5，10，每一折用作验证集，其余各折用作训练集。我们默认使用3折，并将其记为CV K折；在与其他折数进行比较时，则特别使用CV_3折。
- 近似交叉验证（ACV）：我们考虑ACV、ProxACV [46] 和ACV K折，后者是针对大型 n 和 p对ACV的一种启发式扩展。
- （扩展）贝叶斯信息准则（BIC/EBIC）：对数似然通过$\lambda = 0$ 估计，采用零支撑集约束，并使用 R 软件包 [glasso 13], ；EBIC 则依据 [11]中的公式 1，使用 $\gamma = 1$ 。

备注 我们使用 glasso R 包 [13] 估计精度矩阵，代码可从以下网址获取：https://github.com/cran/glasso。我们按照 [46]中的相同构造，使用 QUIC [14] 计算 ProxACV。 λ 值集是预先计算的。 $\lambda_{\max}$ 是每个节点与其余 [12]之间相关性的无穷范数最大值。我们按固定比例逐步减小 $\lambda_{\max}$ ，以找到使 glasso 得到平凡全图估计的 $\lambda_{\min}$ 。时间以对数尺度绘制。
我们注意到，对于每个候选支撑集，BIC/EBIC 都需要通过额外的支撑集约束的无惩罚 GLasso 优化来计算极大似然解。这需要额外的三次计算量来计算高斯似然。相比之下，ACV 使用公式 (4) 中明确给出的单步更新构造每个近似折估计，无需再次优化或运行收敛循环。因此，我们预计这种方法具有计算优势，因为它以固定次数的显式更新取代了 GLasso 中额外的迭代约束求解。

E.2 次高斯模拟

对于次高斯情形，我们考虑高斯分布。具体而言，我们遵循 flare 包[20] ，根据图结构构造均值为 0、协方差矩阵为 n 的高斯分布数据，并将 $p =$ 设为 10 到 5000（10，20，50，100，200，500）。

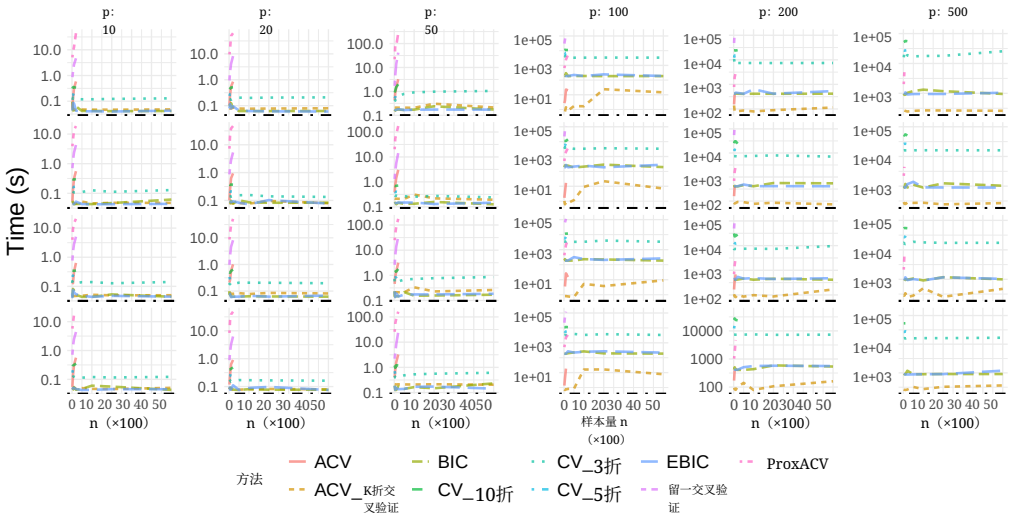


图5：在次高斯分布下的大规模（n 最多为 5000）情形中，通过计算成本评估各方法。LOOCV、ACV 和 ProxACV 的计算开销随 n 扩展，并很快变得不可行。与 CV K-折以及 BIC/EBIC 相比，ACV K-折的计算成本始终低得多。

为了在计算可行的范围内比较所有 CV 变体，我们将 $n \leq 200$ 限制在一定范围内，以展示留一交叉验证及其变体 ACV 和 ProxACV 的优缺点，如图6所示。显然，与 K折交叉验证、贝叶斯信息准则和扩展贝叶斯信息准则相比，留一交叉验证及其变体能更快地收敛至真正率 = 1 。然而，它们的计算成本都随 n扩展，对于 $n \geq 100$很快就会变得不切实际。相比之下，贝叶斯信息准则和扩展贝叶斯信息准则需要更多数据，才能达到与留一交叉验证及其变体相近的真正率水平。不过，它们的计算成本要低得多。

我们在图7和图5中分别展示了 n 最多为 5000 时的综合结果，用于比较扩展至 K-折 CV 的 ACV（ACVK-折）在性能和计算成本方面的表现。图5传达的主要信息是，ACV K-折具有最低的计算成本。此外，图7显示，CV 变体和 CV 近似在预筛选方面表现良好，但由于无论样本量如何增长它们都不一致，因此需要进行后校正。

在固定 $p = 100$ 下预设 α 水平的后校正结果如图8所示。我们选择 n 来概括图估计与真实图之间的超集关系。不过，我们注意到，不同的图集体现了不同的学习难度，因此粗略地设定$\widehat{n} = \{100, 500, 1000\}$，以表示图估计 G的三种情况：

1. 在 $n = 100$ 处不是超集： $G \not\supseteq \widehat{G}$。
2. 在 $n = 500$ 处大致为超集： $G \gtrsim \widehat{G}$。
3. 在 $n = 1000$ 处为超集： $G \supseteq \widehat{G}$。

如第 5 节所述，族错误率可能较宽松；图8显示，错误发现率水平远低于 α阈值化。三个样本选择（$n = 100$，500，1000）对应图估计与真实图之间三种不同的超集关系。当$\widehat{G} \not\subseteq G$ ($n = 100$)时，后校正筛选得过于严格，显著损害了真正率。相比之下，当 $\widehat{G}$ 大致包含真实图 G 时，后校正能够有效去除错误发现，同时保持合理的真正率。相比之下，我们看到，结合后校正的 ACV K 折法在计算成本和结构学习性能两方面均优于贝叶斯信息准则和扩展贝叶斯信息准则。例如，当 $\alpha = 0.5$时，与贝叶斯信息准则和扩展贝叶斯信息准则相比，ACV K 折法的真正率更高，同时错误发现率为 0。

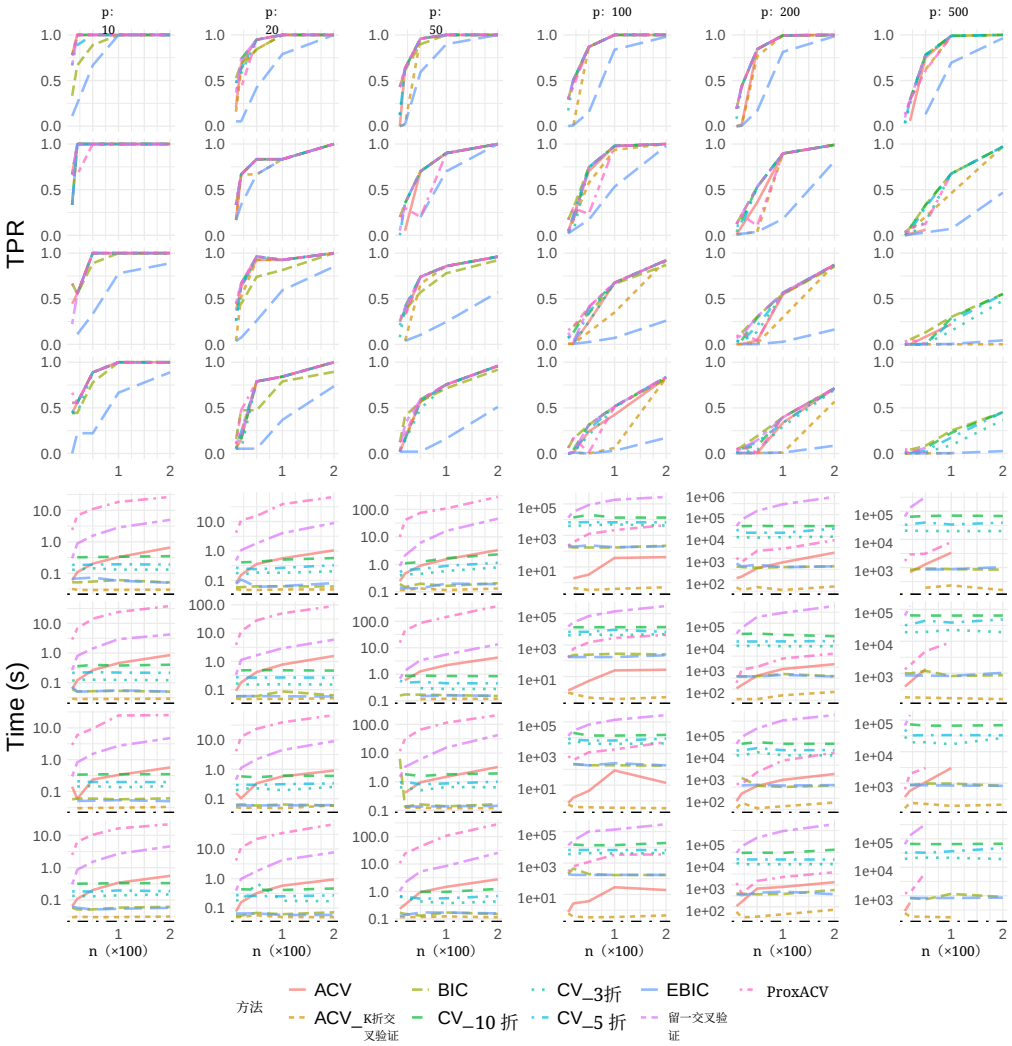


图6：次高斯分布下留一交叉验证变体（ACV 和 ProxACV）与其他方法的比较。留一交叉验证、ACV 和 ProxACV 在基于真正率辨别边方面表现最佳，但它们的计算开销随 n 增长。

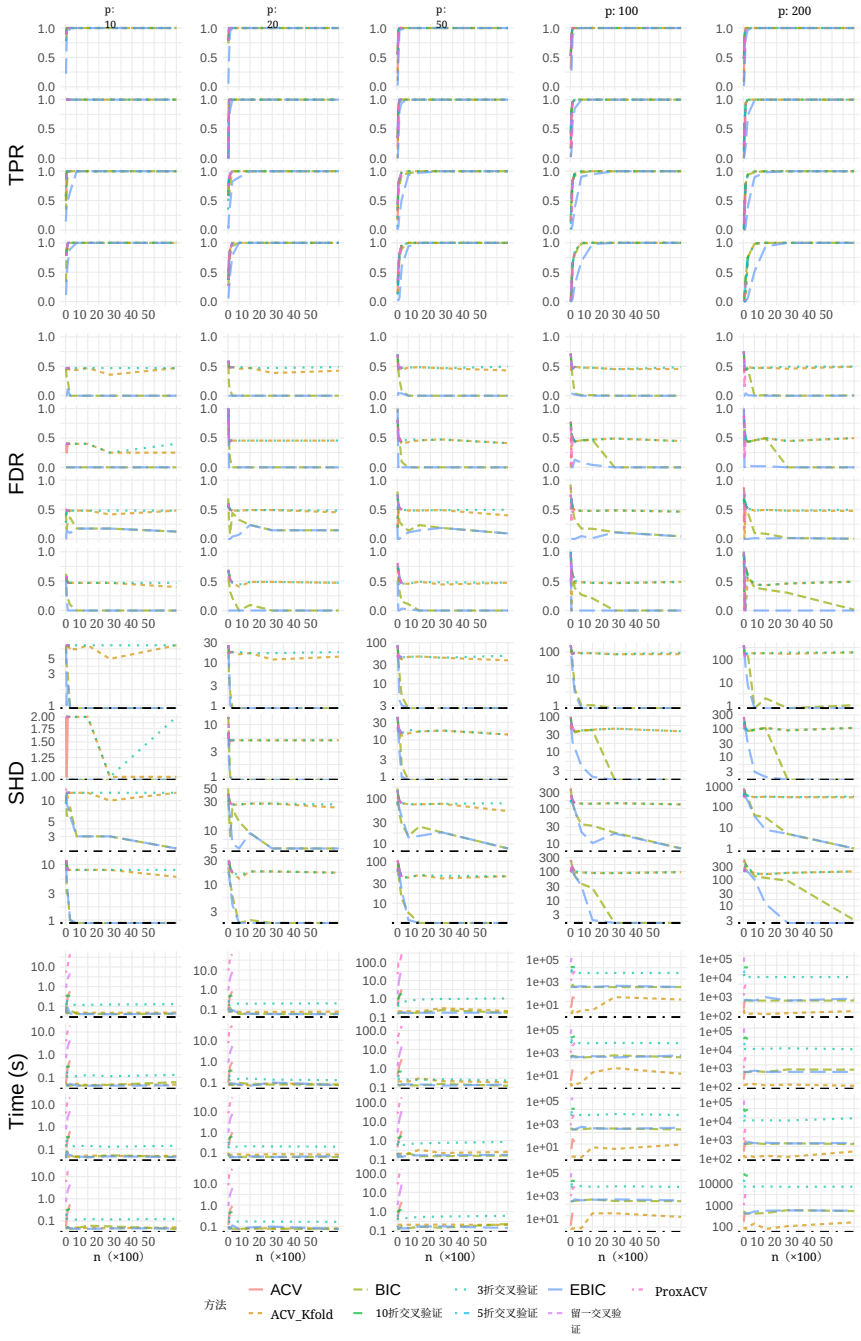


图7：在次高斯分布下，通过大规模评估方法的表现（n 最高达 5000）。在不同样本量 n（以百为单位）和节点数 p 下，针对各类图类型计算 SHD、TPR 和 FDR 的平均值。黑色点划线表示 SHD 为 0，即实现了完美的结构学习。所有 CV 变体和单独使用的近似方法都不一致，FDR 和 SHD 均非零。

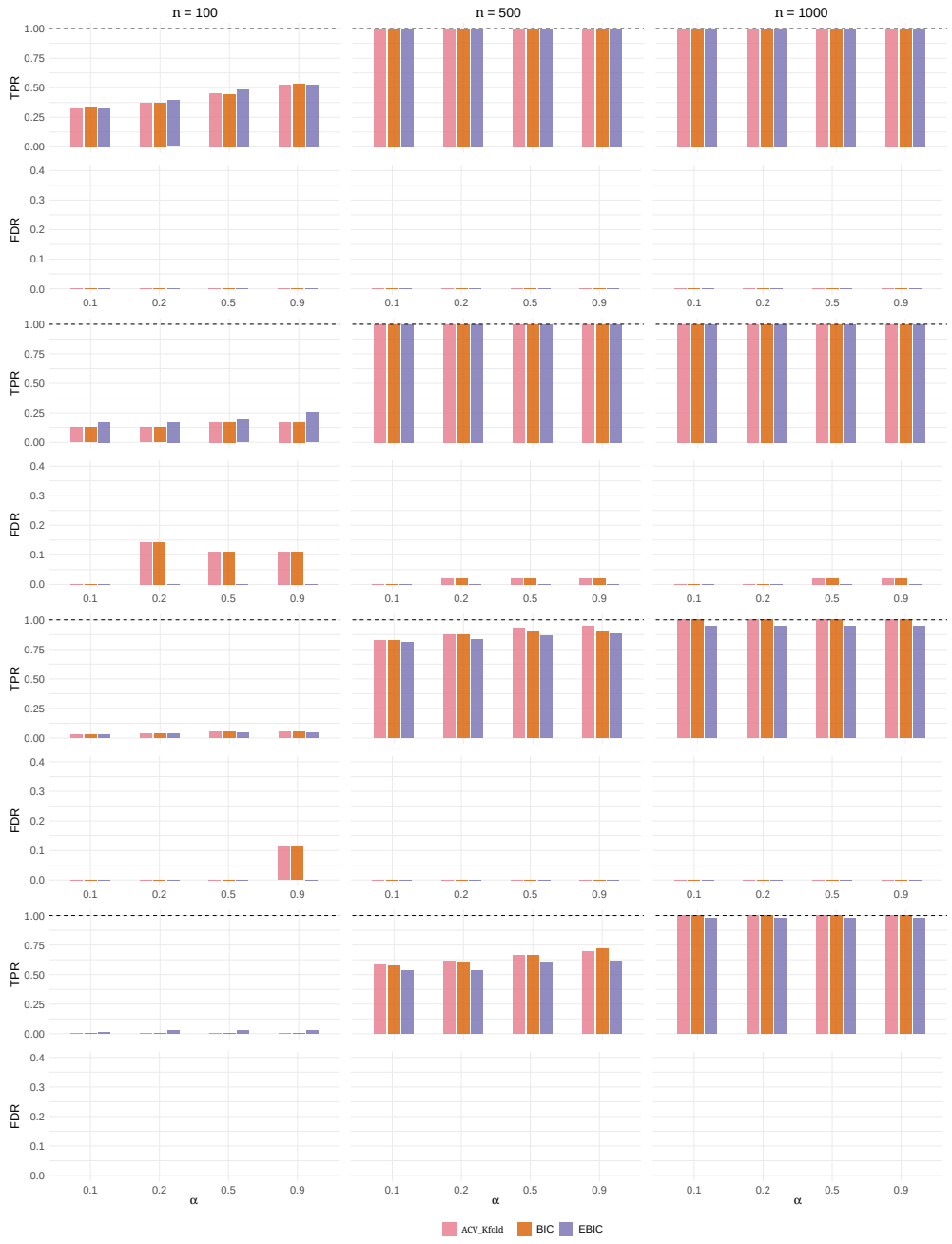


图8：在次高斯分布下，ACV K-折交叉验证、贝叶斯信息准则和扩展贝叶斯信息准则结合后校正后的性能比较，其中 n 变化而 $p = 100$固定。 α 水平为预设值，黑色虚线表示真正率 = 1。

E.3 非次高斯模拟

我们详细介绍用于检验次高斯假设稳健性的对数正态分布情形。我们使用 R 中的 statspackage 构造服从对数正态分布、均值为 0 且标准差为 1 的数据。为便于比较，我们采用与次高斯情形相同的 n， p 设置。

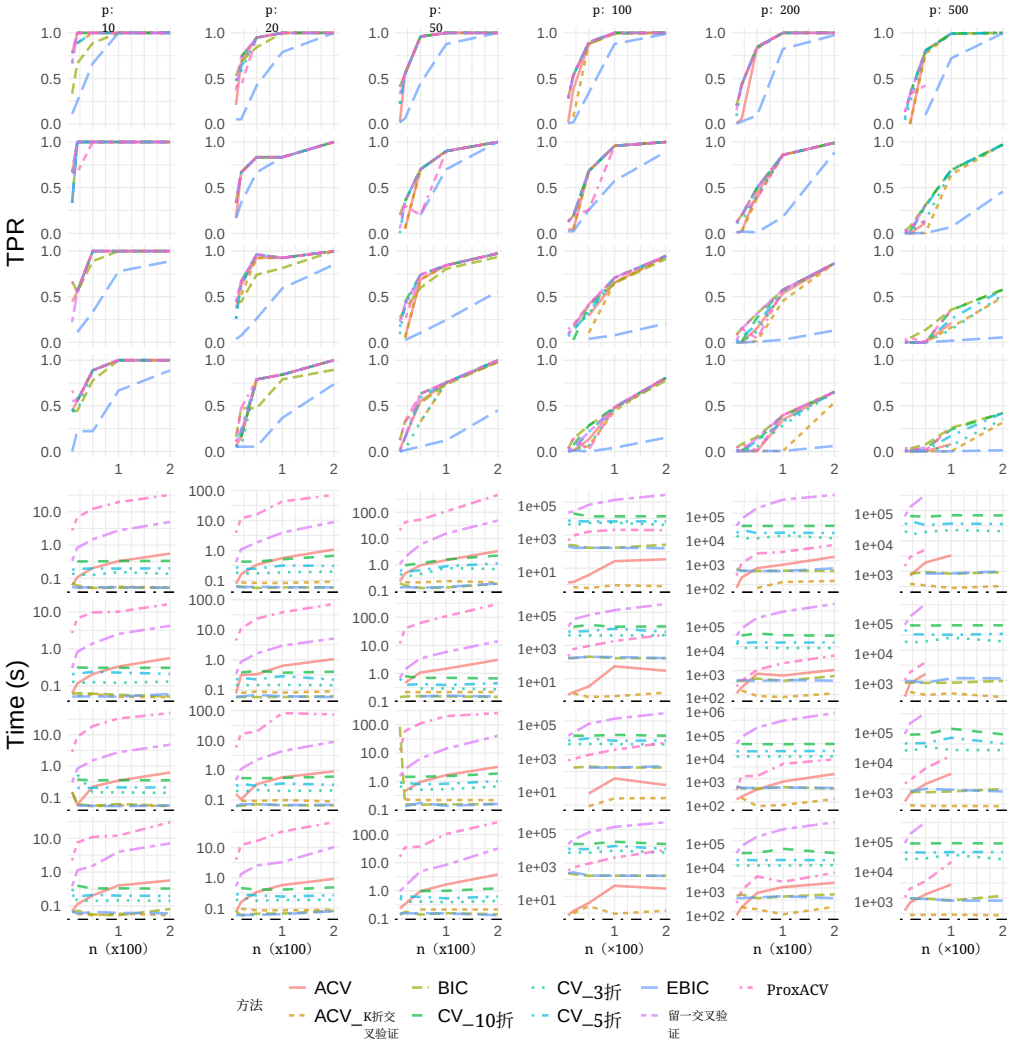


图9：对数正态分布下留一交叉验证变体（ACV 和 ProxACV）与其他方法的比较。根据真正率表现，留一交叉验证、ACV 和 ProxACV 在辨别边方面表现最佳，但其计算开销随 n 增长。时间以对数尺度绘制。

图9显示，尽管留一交叉验证、ACV 和 ProxACV 使用比 K 折交叉验证、贝叶斯信息准则和扩展贝叶斯信息准则更少的样本便能在边检测方面取得出色表现，但它们在计算方面处于劣势。我们还分别在图10和图11中展示了大样本量下的计算成本和结构学习表现。传达的信息与次高斯情形相同：ACV 是一种合格且高效的预筛选器，但要实现一致性仍需进行后校正。

图 14 展示了结合后校正的 ACV K 折、贝叶斯信息准则和扩展贝叶斯信息准则的评估结果。除了在次高斯情形下得出类似结论外，我们还观察到，对于 K近邻和 SF 等稀疏图，当数据不足时，ACV K 折能识别出更多真实边（$n = 500$）。这种继承自交叉验证、能在样本有限时进行边检测的非凡能力，对于实现一致性至关重要。例如，当 $\alpha = 0.5$ 时，ACV K 折对 K近邻和 SF 具有一致性，而贝叶斯信息准则和扩展贝叶斯信息准则则无法找出所有真实边（真正率 $\neq 1$）。

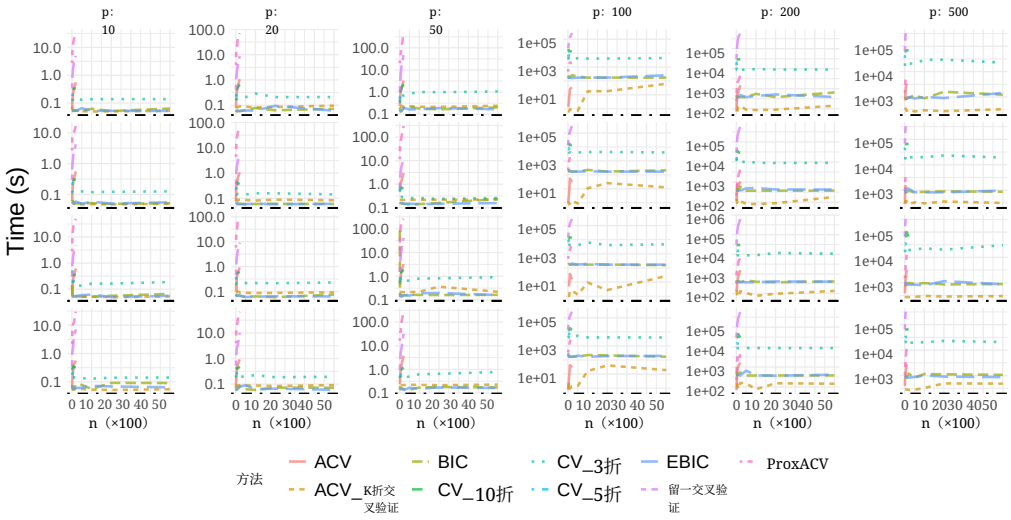


图10：在对数正态分布下的大规模（n 最高达 5000）场景中，通过计算成本评估各方法。LOOCV、ACV 和 ProxACV 的计算开销随 n 扩展，并很快变得不可行。与K-折 CV 以及 BIC/EBIC 相比，ACV K-折的计算成本始终低得多。时间以对数尺度绘制。

E.4 非线性模拟

对于非线性情形，我们考虑伊辛模型。数据的每一行 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 均为独立同分布样本，根据图集通过 IsingSampler[9], 生成；我们通过 [35]中的基于邻域的逻辑回归构造 ℓ_1 惩罚估计量。令 $\lambda > 0$。不失一般性，我们通过对其余节点进行回归，估计第 p 个节点的邻居，具体如下：

$$\widehat{\theta}^{\lambda,\mathbf{X}} := \arg\min_{\theta\in\mathbb{R}^{p-1}} \frac{1}{n}\sum_{i=1}^{n} f(\theta;\mathbf{X}^{(i)}) + \lambda\|\theta\|_1,$$

$$f(\theta;\mathbf{X}^{(i)}) := \log\left(e^{\sum_{j=1}^{p-1}\theta_j\mathbf{X}_j^{(i)}} + e^{-\sum_{j=1}^{p-1}\theta_j\mathbf{X}_j^{(i)}}\right) - \sum_{j=1}^{p-1}\theta_j\mathbf{X}_p^{(i)}\mathbf{X}_j^{(i)}$$

$$\widehat{\mathsf{ne}}^{\lambda,\mathbf{X}} := \{j\in[p-1] \mid \widehat{\theta}_j^{\lambda,\mathbf{X}} \neq 0\}.$$

一旦我们估计出每个节点的邻居 $\widehat{\mathsf{ne}}^{\lambda,\mathbf{X}}$，就可以结合 AND 或或规则，根据 [31] 中的讨论，对估计图进行对称化。近似交叉验证的推导与向量形式类似，详细推导见 [39] 中的附录 D.4。

$$\tilde{\theta}^{\lambda,\mathbf{X}} := \begin{bmatrix} \widehat{\theta}^{\lambda,\mathbf{X}}_{\hat{\mathcal{S}}} + (H^{(-i)}_{\hat{\mathcal{S}}\hat{\mathcal{S}}})^{-1}[\nabla_\theta f(\theta;\mathbf{X}^{(i)})]_{\hat{\mathcal{S}}} \\ \mathbf{0}_{\setminus\hat{\mathcal{S}}} \end{bmatrix}, \quad H^{(-i)}_{\hat{\mathcal{S}}\hat{\mathcal{S}}} = H_{\hat{\mathcal{S}}\hat{\mathcal{S}}} - [\nabla^2_\theta f(\theta;\mathbf{X}^{(i)})]_{\hat{\mathcal{S}}\hat{\mathcal{S}}},$$

其中，$H_{\hat{\mathcal{S}}\hat{\mathcal{S}}}$ 是在支撑集 $\hat{\mathcal{S}} = \mathrm{supp}(\widehat{\theta}^{\lambda,\mathbf{X}})$ 上的受限海森矩阵。

不同于式（4）中的克罗内克积，该近似涉及复杂且开销高昂的海森逆矩阵向量乘积计算。我们采用预条件共轭梯度法来近似 ACV K-折的海森逆矩阵乘积，结果如图 15 所示。我们注意到，还可以通过其他替代方案加以改进；由于这与我们的稳健性测试重点正交，因此将可能的改进留待未来工作。

我们从图15中发现，海森逆矩阵乘积计算量的减少并未损害预筛选能力，因此 ACV K 倍仍是合格的后校正预筛选器。此外，图15显示，随着观测值数量和变量数量不断增大，ACV 的计算优势很快便会显现并持续保持。

后校正结果如图16所示。我们注意到，当图估计在 $n = 100$处根本遗漏了真实边时，后校正无法控制错误发现。相比之下，

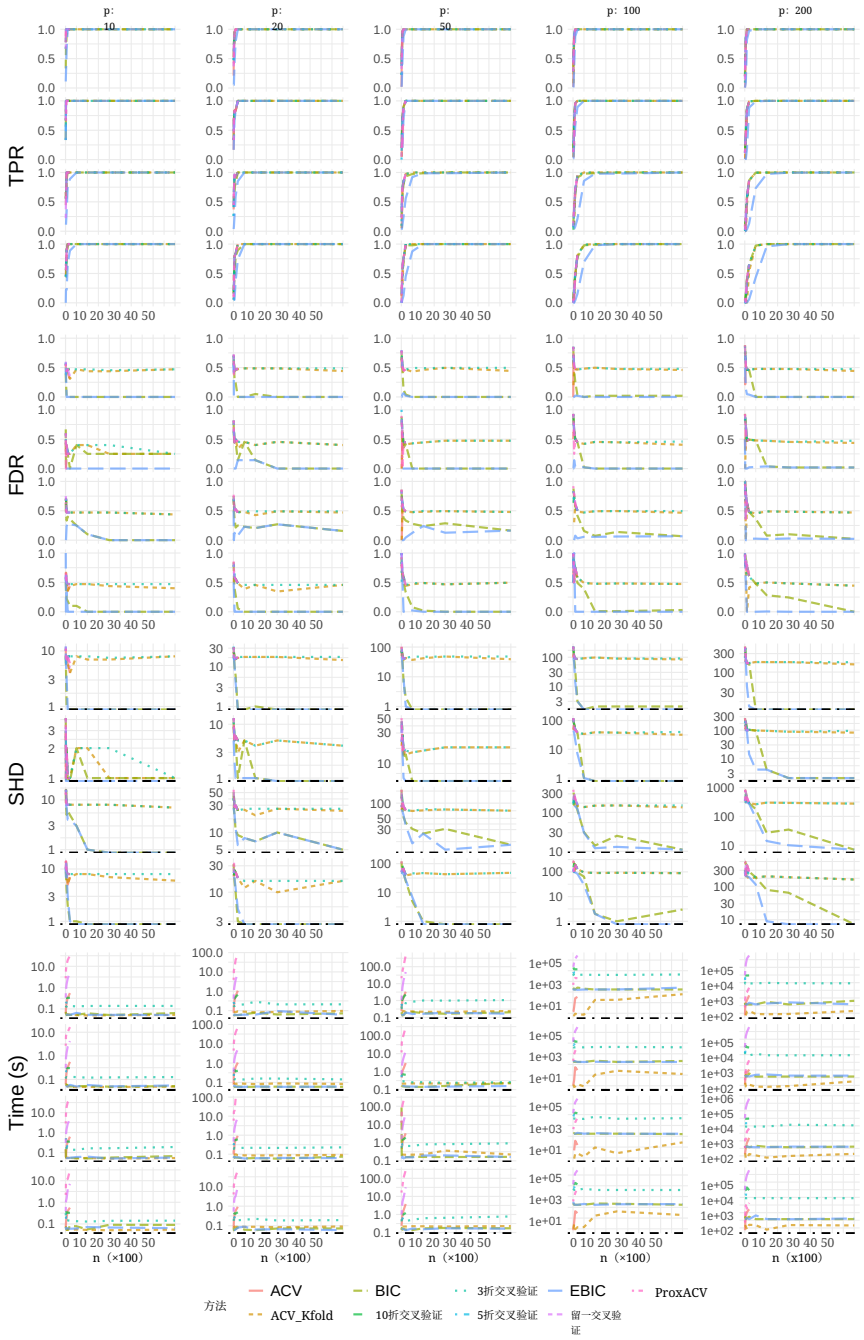


图11：在对数正态分布下，对大规模（n 高达 5000）方法性能的评估。不同样本量 n（以百为单位）和变量数量 p 下，各类图类型组的平均结构汉明距离、真正率和错误发现率。黑色点划线表示结构汉明距离为 0，即实现了完美的结构学习。所有 CV 变体以及单独使用的近似方法，在非零错误发现率和结构汉明距离方面均不一致。

我们在图 8 和图14中发现，当线性条件满足时，即使图估计遗漏了真实边，错误发现率也能得到良好控制。相反，当真实图包含在图估计中时（最后一列），我们看到 ACV 能够成功筛除错误发现，并且对于 $n = 1000$处非线性伊辛模型的带状图、ER图和SF图，结果具有一致性。即使真实图在 $n = 500$处大致包含在图估计中，我们也观察到所有情况下的错误发现率都得到了良好控制，即使线性

假设不成立。简言之，当图估计包含或大致包含真实图时，与贝叶斯信息准则和扩展贝叶斯信息准则相比，ACV 的预筛选效果更好，计算成本更低。结合后校正后，ACV 的表现优于贝叶斯信息准则和扩展贝叶斯信息准则。

E.5 实践中调节 α

在实践中，真实图是未知的，选择合适的 α 颇具挑战。一方面，较小的 α 过于严格，尤其是在 $G \not\supseteq G$时，即图 2 所示的非超集情形。另一方面，较大的 α 无法达到控制错误发现的目的，例如图 16中 $n = 1000$ 对应的 $\alpha = 0.9$ 。

直观而言，模型选择比预测任务更进一步；预测性能良好的模型未必是真实模型，但真实模型理应总能有良好的预测性能。附录 F 中的 [39]充分体现了这一点。换言之，如果较小的 α 能够达到与其他较大的 α相近的性能，那么较小的 α 更可取。因此，我们可以通过比较估计图 G^{α} 的预测损失与其他 α 选项对应的预测损失，来评估 α 。给定 $\widehat{G}^{\lambda}$ 和 $\alpha \in (0,1)$，我们进行后校正，得到估计图$\check{G}^{\alpha}$，随后重新拟合该估计图，并计算其相对于 $\widehat{G}^{\lambda}$所对应误差的相对误差。我们以 $p = 100$ 情形下的 $\widehat{G}^{\lambda_{\mathrm{ACV}}}$ 为例（图 2 中由 ACV K折得到的估计图），展示不同$\alpha \in [0.1, 0.9]$下的结果。我们还纳入 $n \in \{100, 500, 1000\}$，以对应$\widehat{G}^{\lambda_{\mathrm{ACV}}}$ 与 G之间的三种超集关系。

我们观察到，当图估计不是真实图的超集（$n = 100$）时，相对误差极小，但没有规律性的趋势。这提醒我们，所用图估计的质量会影响后校正，因为包含所有真实边的图不应导致预测性能出现反复波动。相比之下，在另外两种情况中，我们观察到了规律性的下降趋势，即 $G \supsetneq \widehat{G}$ 和 $G \supseteq G$ （$n = 500,\ 1000$）。当估计图大致是一个超集（第二列）时，即使 n 不断增加，小 α 和大 α 之间的差距依然存在。这表明，持续放宽对错误发现的限制会带来性能提升，因此，如果我们的目标是预测任务，就应采用限制较少的 α 。这一观察与图2第二列的结果一致：随着 α 增加，真正率也随之上升（但未达到1）。相比之下，当估计图是真实图的超集（最后一列）时，这种差距会逐渐缩小。这表明，放宽对错误发现的限制只能带来微乎其微的提升。因此，我们可以预期更小的

α 在性能大致相当的情况下，具有更好的错误发现控制。

我们注意到，调节 α 相比 λ有两项优势。首先， α 直接编码目标错误率（例如，错误发现率水平），而 λ 是正则化路径上的尺度参数，无法明确对应到所需的错误率。其次， α 的取值范围固定且易于解释，为 $(0,1)$，不受数据尺度或模型维度影响；而 λ 的适当取值范围取决于具体问题，通常通过启发式方法选定。给定 α，我们的方法通过 ACV 自动选择 λ，然后在后校正步骤中应用α水平的偏相关检验。

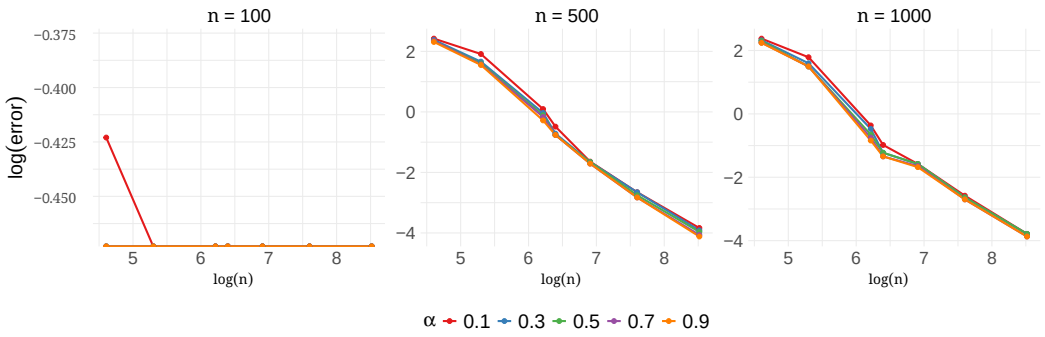


图 12：通过预测误差评估 α，并考察其与样本量 n 的关系。

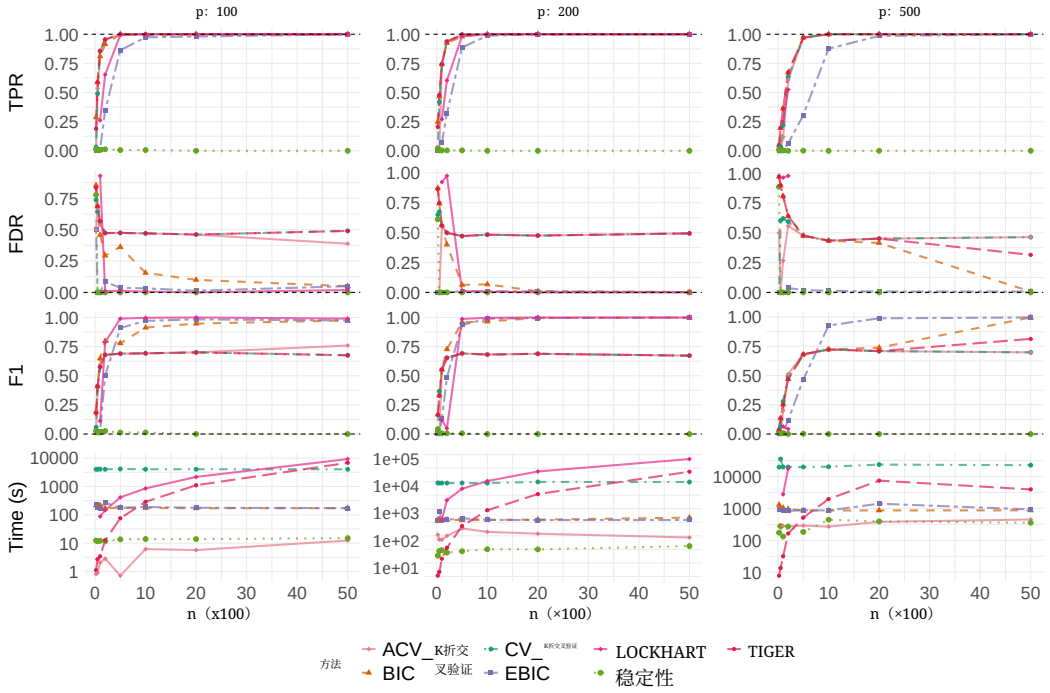


图 13：高斯分布下各方法在预筛选性能方面的比较。就预筛选而言，ACV K 折交叉验证在计算成本和所需样本数方面均优于其他方法。

表 2：后校正后的结构学习之间的 F1 分数比较。

α	ACV K-折	BIC	EBIC	交叉验证 K-折	LOCKHART协方差检验	稳定性	TIGER
0.1	0.977	0.977	0.915	0.977	0.977	0.013	0.977
0.3	0.990	0.987	0.919	0.990	0.987	0.013	0.990
0.5	0.997	0.990	0.923	0.997	0.990	0.013	0.997
0.7	0.994	0.997	0.923	0.994	0.990	0.013	0.994
0.9	0.990	0.993	0.923	0.990	0.990	0.013	0.990

E.6 与现有结构学习方法的比较

我们在第 6.1 节的相同预筛选设置下，加入稳定性选择 [24], TIGER [22] 和 Lockhart协方差检验 [25]，以聚焦于高斯数据下的 K 近邻图。我们的实现基于 huge 和 flare R 包 [20, 50]。我们还加入 F1 分数作为综合度量，以便进行更直观的比较。预筛选结果如图 13 所示。我们发现，虽然稳定性选择的计算量最小，但几乎没有检测到任何有意义的边。此外，我们注意到，当 $p = 500$时，EBIC 的 F1 分数优于 ACV K折；但要达到真正率 = 1，它所需的数据量远多于 ACV K折。当 $n < p$时，ACV K折以最小的计算量实现了最佳边检测效果，显示出其在高维场景下出色的预筛选能力。

为了更好地比较后校正后的整体表现，我们关注 $p = 100$ 和$n = 500$ 在F1 分数上的表现，结果见表2。首先可以看到，随着错误发现率控制减弱（ α增加），稳定性方法的F1 分数保持不变。这表明其预筛选存在问题。我们还注意到，扩展贝叶斯信息准则和Lockhart协方差检验的F1 分数随着 α逐渐提高，体现了对错误发现进行控制的努力。相比之下，其他方法的F1 分数均先升后降。这意味着在某个 α 水平下，所有真正例均已纳入，而进一步增加 α 会引入错误发现，导致F1 分数降低。对此还有一种解读：只有ACV K折、3 折交叉验证和TIGER

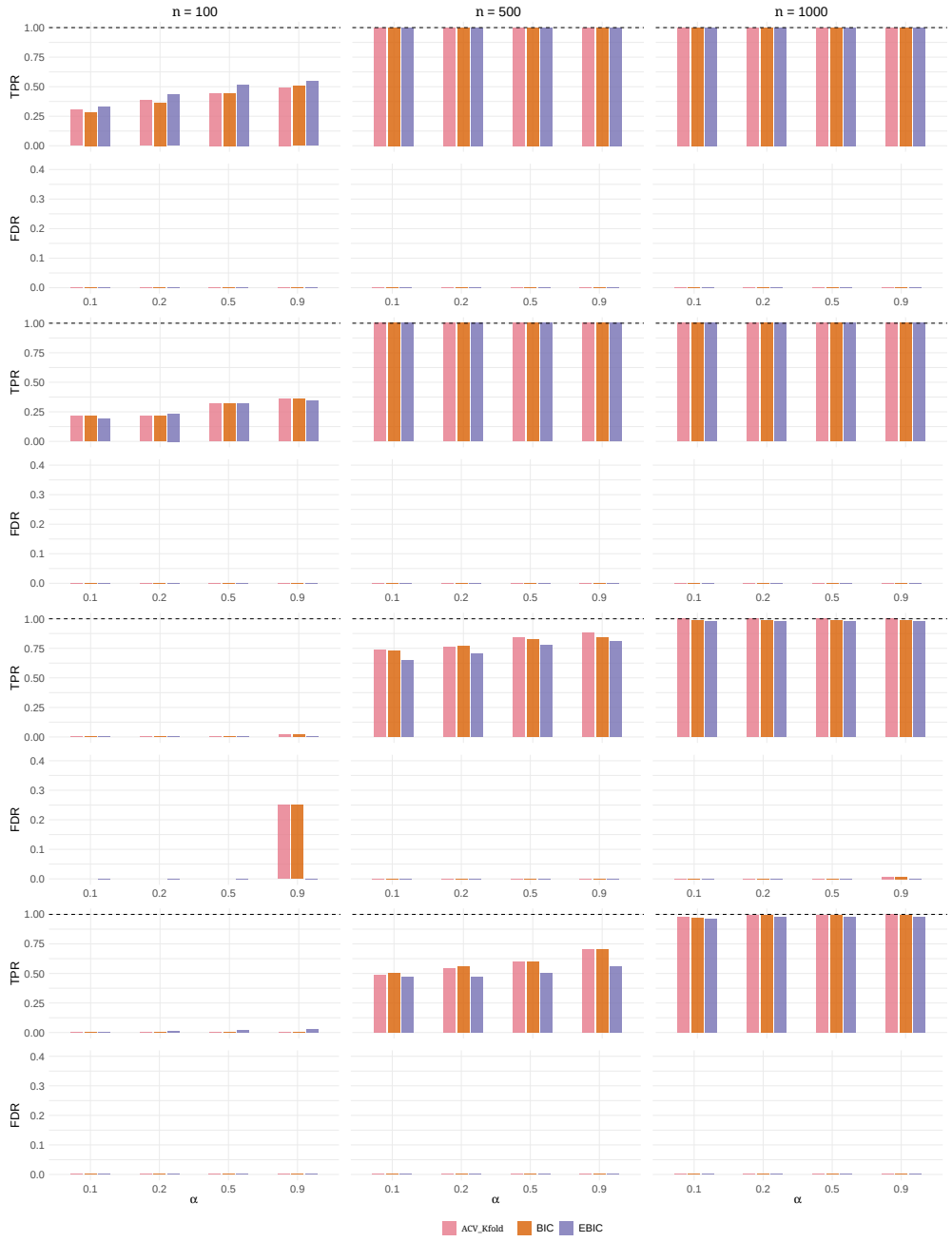


图14：在对数正态分布下，比较结合后校正的 ACV K-折交叉验证、贝叶斯信息准则和扩展贝叶斯信息准则在 n 变化、$p = 100$ 固定时的性能。预先设定 α 水平，黑色虚线表示真正率 = 1。

以最小的 α达到了最高F1 分数；在这些方法中，ACV K折的计算效率最高。

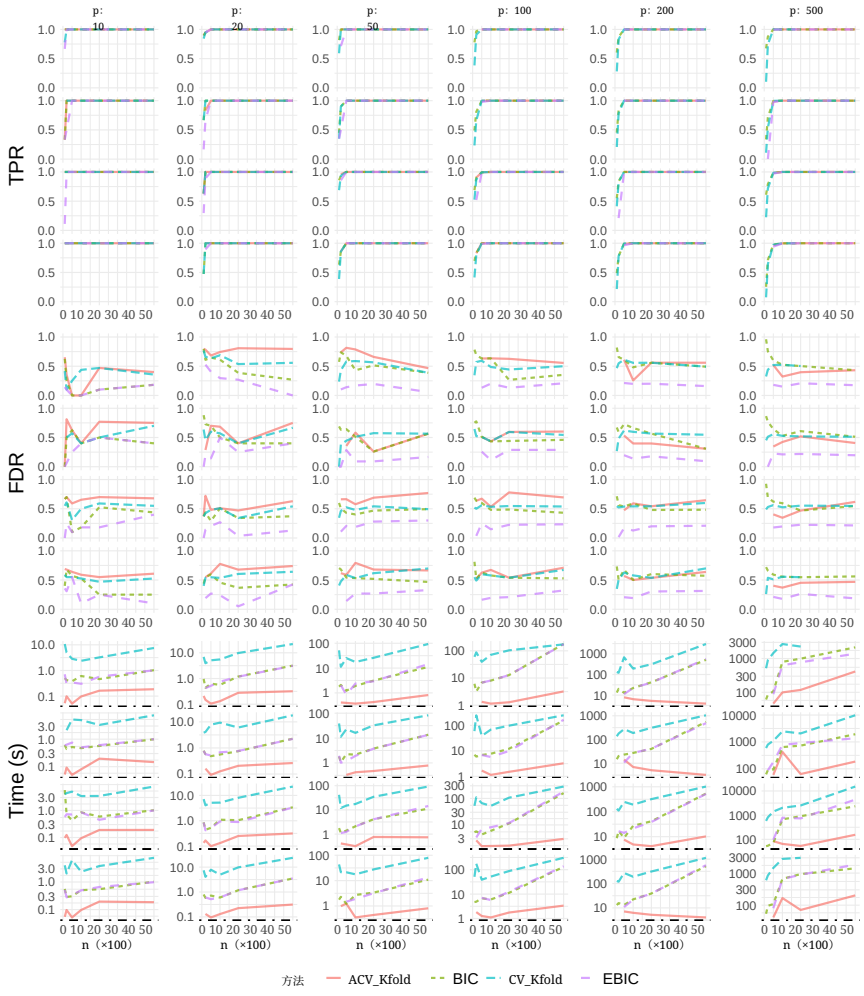


图15：在伊辛模型下，通过大规模性能评估（n 最高可达 5000）。在不同样本量 n（以百为单位）和变量数 p 下，按图类型分组，计算真正率、错误发现率和时间的对数尺度平均值。



图16：在伊辛模型下，ACV K 折、贝叶斯信息准则和扩展贝叶斯信息准则结合后校正的性能比较，其中 n 变化而 $p = 100$ 固定。 α 水平预先设定，黑色虚线表示真正率 = 1。