arXiv:2302.05787v1 [stat.ML] 11 Feb 2023

用于密度估计、数据合成和变分推断的差分隐私归一化流：在电子健康记录中的应用

Bingyue Su*、Yu Wang*、Daniele E. Schiavazzi 和 Fang Liu

圣母大学应用与计算数学及统计学系，美国印第安纳州圣母城

摘要

电子健康记录（EHR）通常包含个体患者的敏感医疗信息，这在共享或发布电子健康记录数据以用于下游学习与推断任务方面造成了显著限制。我们采用一类深度生成模型——归一化流（NF），在提供差分隐私（DP）保证的前提下估计数据集的概率密度，并据此生成隐私保护合成数据。我们将该技术应用于一个包含肺动脉高压患者的电子健康记录数据集。通过比较高血压状态预测的准确性以及基于物理的模型参数的变分后验分布，评估合成数据在学习与推断方面的效用。此外，我们使用一个非线性模型生成的模拟数据集，比较基于隐私保护合成数据的变分推断（VI）结果，以及在给定原始非隐私数据集的情况下，通过直接对用于变分推断的归一化流实施差分隐私而获得的隐私保护变分推断结果。结果表明，通过采用归一化流进行差分隐私密度估计所生成的合成数据，能够以合理的隐私成本实现良好的效用。我们还表明，基于自由能界损失、通过差分隐私归一化流获得的变分推断，可能会产生相关结构显著改变的变分近似；而基于两种概率分布间其他差异度量的损失形式，或可改善结果。

关键词：差分隐私、隐私保护、合成数据、归一化流、变分推断、自由能界、KL散度、密度估计、电子健康记录

1 引言

1.1 背景

过去十年来，随着数据体积呈指数级增长、强大高效的数据分析技术迅速发展，以及计算设施和

*共同第一作者

资源日益普及，数据科学作为一门跨学科领域获得了巨大关注。在研究人员之间共享数据或向公众发布信息时，始终存在隐私问题，以及重新识别或泄露敏感信息的风险。即使移除了关键标识符，对手仍可能利用公开可用的数据和复杂的攻击手段识别某个人，或获知其敏感信息。一些值得注意的隐私攻击包括基于健康记录的重新识别（Sweeney，1997，2013年）、Netflix Prize（Narayanan and Shmatikov，2008年）、成员关系攻击（Backesetal.，2016年；Homeretal.，2008年；Shokrietal.，2017年）等。

在本研究中，我们以电子健康记录（EHR）为例，介绍隐私保护数据共享与分析的方法。电子健康记录数据通常包含个体患者的敏感医疗信息。例如，心血管电子健康记录是本研究考察的电子健康记录数据类型，其中包含心率、动脉和静脉压力、体循环和肺循环压力、心输出量、心房和心室压力与容积、射血分数等生命体征测量值，这些数据可用于诊断高血压。由于隐私问题，出于研究目的访问大型单中心电子健康记录数据集通常存在困难（参见近期将大型临床数据集公开的努力，例如Johnson 等人，2020年中讨论的 MIMIC-IV 数据集），更不用说合并多个电子健康记录数据源，以创建能够训练复杂预测模型的大型多中心数据集了（在此背景下，可参见Pollard 等人，2018年中讨论的 eICU Collaborative Research Database）。

为减轻发布信息和共享数据时的隐私问题，过去二十年间人们提出了各种隐私概念，例如 k-匿名性（Samarati 和 Sweeney，1998年）、 t-接近性（Li 等人，2007年）、 l-多样性（Aggarwal 和 Yu，2008年）以及差分隐私（DP）（Dwork 等人，2006a，b，c）等。我们重点关注 DP，这是当代隐私研究中一种广受欢迎的隐私概念和框架。DP 不仅能够为隐私提供数学保证，并且能够抵御各种隐私攻击（Dwork 等人，2017年），还具备隐私损失组合和不受后处理影响等优良性质，这些性质显著促进了其在研究和实际应用中的普及。针对广泛的统计方法和机器学习方法，已有具备差分隐私保证的隐私保护技术和算法，涵盖从发布计数和直方图，到发布回归模型或深度神经网络的预测结果。

在本研究中，我们考察差分隐私归一化流（NF），分别通过密度估计生成合成数据，并在预先设定的隐私预算下获得具备差分隐私保证的变分推断（VI）。NF 是生成模型（Dinh 等人，2016年；Kingma 等人，2016年；Papamakarios 等人，2017年；Rezende 和 Mohamed，2015年），主要应用于密度估计和变分推断。

1.2 我们的工作与贡献

我们的工作分为两部分。第一部分重点研究如何通过差分隐私密度估计生成合成数据，以及合成数据在下游学习任务中的效用。据我们所知，只有另一篇论文（Waites 和 Cum- mings，2021）探讨了在用于密度估计的正规化流中引入差分隐私。此外，此前没有研究探讨使用通过 DP-NF 生成的合成数据进行预测或推断等下游学习任务。我们的工作旨在填补这一空白。

具体　此外，我们将 DP-NF 应用于包含　s

一组患者的血流动力学测量数据，其中包括肺动脉压力升高者（II 组肺动脉高压，参见 Simonneau 等人，2013年）和射血分数保留型心力衰竭患者（Harrod 等人，2021）。该电子健康记录数据集的样本量较小，属性数量相对较多，包含连续或离散的数值测量值以及缺失值。这为考察 DP-NF 在数据类型异质性条件下，为小型数据集生成实用且保护隐私的合成数据的可行性提供了机会；与模拟数据或经过充分研究的基准数据集相比，此类数据也具有更高的真实度。我们使用 DP-NF 生成的合成数据进行回归和分类，并将预测准确率与使用原始数据所获得的结果进行比较。我们还利用合成数据，对一个计算成本高昂的基于物理的模型进行变分推断。为降低计算成本，我们采用了一个针对该基于物理的模型离线训练的快速神经网络代理模型。

本文的第二部分探讨了隐私保护变分推断（VI）：为变分推断直接对正规化流（NF）进行隐私化，并提供差分隐私保证（而非利用正规化流估计的差分隐私密度所生成的隐私合成数据进行变分推断）。我们在模拟数据上考察了该方法推断非线性回归模型参数的性能。据我们所知，本文首次探讨了使用自由能界损失和基于随机梯度的优化，对用于变分推断的正规化流进行差分隐私化的可行性。我们表明，这种损失函数形式可能会得到与目标密度显著偏离的差分隐私变分分布；特别是当目标分布为多元分布时，偏差会体现在相关性上，因为与目标分布相比，差分隐私保证会使差分隐私变分分布中的边际方差膨胀。我们还考察了若干其他用于 NF-VI 的差异度量损失函数。其中一些度量呈现出与自由能界损失相似的现象，但并非全部如此。这一现象说明，基于 DP-NF 进行变分推断时，仔细选择损失函数至关重要。

在这两部分中，我们都通过随机梯度下降优化应用高斯差分隐私来实现隐私保证，这与现有研究中用于跟踪隐私损失的矩会计方法（Abadi等人，2016年）形成对比（Waites 和 Cummings，2021）。与基于 Rényi 差分隐私的矩会计技术相比，高斯差分隐私对隐私损失组合给出了更紧的界，并且被证明能够在深度学习中实现更高的预测准确率（参见Bu 等人，2020年），而矩会计技术基于 R´ényi 差分隐私。

1.3 相关工作

一种直接实现差分隐私密度估计并生成合成数据的方法，是采用差分隐私进行直方图清理，例如扰动直方图或平滑直方图方法（Wasserman 和 Zhou，2010年）。直方图作为密度估计器会受到维度灾难的影响，而加入差分隐私后，这一问题会更加严重；此外，使用直方图估计器时还需要对连续属性进行离散化。Hall 等人（2013年）和 Alda 和 Rubinstein （2017年）开发了用于保护函数输出隐私的差分隐私机制，并将其应用于核密度估计。研究人员还提出了将局部差分隐私用于非参数密度估计的方法（Butucea 等人，2020年；Kroll，2021）。Waites 和 Cummings（2021）通过采用矩会计方法（Abadi 等人，2016年）迭代地对随机梯度下降进行隐私处理，以追踪隐私损失，从而开发了用于密度估计的隐私保护正规化流。

关于具有隐私保证的变分推断，也已有相关研究。Karwa 等人（2015年）直接对用于净化似然函数的差分隐私机制建模，并对由此得到的经隐私处理的后验分布进行变分近似。J¨alk¨o 等人（2016年）在假设变分分布族的前提下，采用矩会计方法对随机梯度下降进行隐私处理。Park 等人（2016年）提出了隐私保护变分贝叶斯框架：当贝叶斯模型属于共轭指数族时，使用高斯机制对潜变量分布上的期望充分统计量进行净化。Sharma 等人（2019年）研究了联邦学习环境下的差分隐私变分推断。J¨alk¨o 等人（2022年）提出了对齐梯度，以考虑与多维变分参数相关的梯度向量中不同元素往往具有不同量级这一事实，旨在不增加额外隐私成本的情况下提高差分隐私变分推断的效用。据我们所知，以上工作均未通过正规化流实现差分隐私变分推断。

用于统计披露限制的数据合成最早由 Rubin（1993年）提出，政府机构利用该方法发布调查数据，医疗保健行业则利用该方法共享医疗数据，此外还应用于其他领域。合成的替代数据集与原始数据集具有相同的结构，但由伪个体组成。差分隐私数据合成（DIPS）为将形式化的隐私保证融入数据合成提供了一种解决方案。DIPS 可通过无模型方法和基于模型的方法实现；有兴趣的读者可参阅 Bowen 和 Liu（2020年），了解近期部分 DIPS 技术的概述。变分自编码器（VAE；Kingma 和 Welling，2013年）、生成对抗网络（GANs；Goodfellow 等人，2014年）以及正规化流（NF；Papamakarios 等，2017年；Rezende 和 Mohamed，2015年）的引入，为通过基于深度神经网络的生成模型进行数据合成带来了新的可能性，并提供差分隐私保证。Xie 等人（2018年）通过生成对抗网络生成隐私保护合成数据（DP-GANs），并利用矩会计方法纳入差分隐私保证。Jordon 等人（2018年）将生成对抗网络与教师集成的私有聚合（PATE）框架（Papernot 等人，2016年、2018年）相结合，以生成差分隐私合成数据，并在实证研究中证明其效用优于 DP-GANs。Chen 等人（2018年）提出了一种基于自编码器的差分隐私生成模型和一种基于变分自编码器的差分隐私生成模型，并证明两种方法都能取得令人满意的效用，但前者相比后者对某些隐私攻击具有更强的鲁棒性。Beaulieu-Jones 等人（2019年）将 DP-GANs 应用于真实临床数据集，以合理的差分隐私保证发布合成数据，并证明了合成数据的效用。Pfitzner 和 Arnrich（2022年）在差分隐私联邦学习中，仅聚合变分自编码器的解码器组件来生成隐私保护合成数据，同时保留编码器在本地服务器上的个性化。

本文其余部分安排如下。第2 节介绍用于密度估计和变分推断的正规化流定义，以及差分隐私的概念。第3 节介绍用于密度估计的差分隐私正规化流流程，以及基于估计密度生成隐私保护合成数据的方法，并将其应用于一个真实的电子健康记录数据集。第4节介绍用于变分推断的差分隐私正规化流流程，并将其应用于由非线性模型模拟的重复测量数据；此外，还深入分析了该流程得到的变分分布的效用。本文最后在第5节总结全文并提出一些结语。

2 预备知识

2.1 归一化流

归一化流（NF）定义为一个映射 F：$\mathbb{R}^d \times \Lambda \to \mathbb{R}^d$，由 $\lambda \in \Lambda$ 参数化，将易于采样的基分布（例如 $\boldsymbol{z}_0 \sim \mathcal{N}(0, I_d)$）中的样本变换为来自所需目标密度的样本。具体而言，F 由一系列复合的 K 双射 F_k：$\mathbb{R}^d \times \Lambda_k \to \mathbb{R}^d$ 构成，每个双射都由 $\lambda_k \in \Lambda_k$：$F_\lambda(\boldsymbol{z}_0) = F(\boldsymbol{z}_0; \lambda) = [F_K(\cdot; \lambda_K) \circ F_{K-1}(\cdot; \lambda_{K-1}) \circ \cdots \circ F_1(\cdot; \lambda_1)](\boldsymbol{z}_0)$ 参数化，其中 $\boldsymbol{z}_k = F_k(\boldsymbol{z}_{k-1}; \lambda_k)$ 对于 $k = 1, \ldots, K$。由于 $F_k(\cdot; \lambda_k)$ 是从 $\boldsymbol{z}_{k-1}$ 到 $\boldsymbol{z}_k$ 的双射，$q_k(\boldsymbol{z}_k)$，因此可通过变量替换得到 $\boldsymbol{z}_k$ 的分布

$$q_k(\boldsymbol{z}_k) = q_{k-1}(\boldsymbol{z}_{k-1}) \left| \det \frac{\partial F_k^{-1}}{\partial \boldsymbol{z}_{k-1}} \right| = q_{k-1}(\boldsymbol{z}_{k-1}) \left| \det \frac{\partial F_k}{\partial \boldsymbol{z}_{k-1}} \right|^{-1}. \tag{1}$$

取对数并对 k求和后，式（1）变为

$$\log q_K(\boldsymbol{z}_K) = \log q_0(\boldsymbol{z}_0) - \sum_{k=1}^{K} \log \left| \det \frac{\partial F_k}{\partial \boldsymbol{z}_{k-1}} \right|. \tag{2}$$

通常，训练 NF 的目标是确定一组最优参数λ，使得 q_K 能够近似目标密度 p（VI）或观测数据 $\boldsymbol{x}$ 的分布（密度估计）。

对于一般的双射，计算式（1）中的雅可比行列式的计算成本可能会随着维度 d 和层数 K 的增加而显著提高。为了高效计算这些行列式，人们提出了多种 NF 架构，分别基于 RealNVP（Dinh 等人，2016年）和 GLOW（Kingma 和 Dhariwal，2018年）等采用耦合层的方法，以及 MAF（Papamakarios 等人，2017年）和 IAF（Kingma 等人，2016年）等采用自回归变换的方法。这两类方法都具有易于计算的三角雅可比矩阵。

在第 3 节和第 4 节将正规化流应用于密度估计和变分推断时，我们使用 MAF（Papamakarios 等，2017）。MAF 的自回归性质通过设置 $p(z_i|z_1, \ldots, z_{i-1}) = \phi((z_i - \mu_i)/e^{\alpha_i})$ 获得，其中 ϕ 是标准正态分布的密度函数，$\mu_i = f_{\mu_i}(z_1, \ldots, z_{i-1})$、$\alpha_i = f_{\alpha_i}(z_1, \ldots, z_{i-1})$，而 f_{μ_i} 和 f_{α_i} 是掩码自编码器神经网络（MADE，Germain 等，2015）。设 $\boldsymbol{z}$ 和 $\hat{\boldsymbol{z}}$ 分别为一个 MADE 网络的输入和输出，该网络具有 L 个隐藏层，每层包含 d_l、$l = 1$、$\cdots$、L 个神经元。层之间的映射可表示为

$$\begin{cases} \boldsymbol{y}_1 = h_1(b_1 + (\boldsymbol{W}^1 \odot \boldsymbol{M}^1)\boldsymbol{z}), & \text{where } \boldsymbol{M}^1 \text{ is } d_1 \times d \text{ and } \boldsymbol{M}^1_{u,v} = \mathbb{1}_{m^1(u) \geq v} \\ \boldsymbol{y}_l = h_l(b_l + (\boldsymbol{W}^l \odot \boldsymbol{M}^l)\boldsymbol{y}_{l-1}), & \text{where } \boldsymbol{M}^l \text{ is } d_l \times d_{l-1} \text{ and } \boldsymbol{M}^l_{u,v} = \mathbb{1}_{m^l(u) \geq m^{l-1}(v)} \\ \hat{\boldsymbol{z}} = h_{L+1}(b_{L+1} + (\boldsymbol{W}^{L+1} \odot \boldsymbol{M}^{L+1})\boldsymbol{y}_L) & \text{where } \boldsymbol{M}^{L+1} \text{ is } d \times d_L \text{ and } \boldsymbol{M}^{L+1}_{u,v} = \mathbb{1}_{u > m^L(v)}, \end{cases}$$

其中，h_l 是层 $l-1$ 和 l 之间的激活函数，对于 $l = 1$，$\cdots$，$L+1$，$m^l(k)$ 是从 1 到 $d-1$ 预先设定或随机生成的整数，b_l 和 $\boldsymbol{W}^l$ 分别是层 $l = 1, \ldots, L+1$ 的偏置和权重参数。

矩阵乘积 $\prod_{l=L+1}^{1} \boldsymbol{M}^l$ 是严格下三角矩阵，满足自回归性质（Germain 等人，2015年）。掩码 $\boldsymbol{M}^1$，……，$\boldsymbol{M}^{L+1}$ 可在一次前向传播中计算所有 μ_i 和 α_i（Papamakarios 等人，2017年）。此外，每个 MAF 层的雅可比矩阵均为下三角矩阵，其行列式等于

| 行列式 $\partial f/\partial \boldsymbol{z}|^{-1}$ = 指数($\sum_{i=1}^{d}\alpha_i$)。在每个 MADE 层之前，可以使用批归一化（Ioffe 和 Szegedy，2015年）对前一层的输出进行归一化。批归一化执行逐元素缩放和平移操作，易于求逆，具有易于计算的雅可比矩阵，并且有助于稳定和加速训练、提高准确率（Papamakarios 等人，2017年）。

使用正规化流（NF）根据观测数据 $\boldsymbol{x}$进行密度估计时，假设有 n 个相互独立的观测值 $\boldsymbol{x}_i$, $i=1$, ..., n，并最大化以下对数似然。

$$\begin{aligned}\ell(\lambda;\boldsymbol{x}) &= \log q_K(\boldsymbol{x}) = \textstyle\sum_{i=1}^{n}\log q_K(\boldsymbol{x}_i)\\ &= \sum_{i=1}^{n}\log q_0(\boldsymbol{z}_{i,0}) - \sum_{i=1}^{n}\sum_{k=1}^{K}\left|\det\frac{\partial F_k}{\partial \boldsymbol{z}_{i,k-1}}\right|,\end{aligned} \tag{3}$$

其中 $\boldsymbol{z}_{i,\ k-1} = F_k^{-1}\circ\cdots\circ F_{K-1}^{-1}\circ F_K^{-1}(\boldsymbol{x}_i)$。通过最大似然方法获得正规化流（NF）参数后，可以先从基分布中采样，再通过一系列双射变换这些样本，从而生成 q_K 的样本。研究表明，正规化流（NF）能够准确逼近各种数据类型的复杂密度函数（Dinh 等人，2016年；Kingma 和 Dhariwal，2018年；Papamakarios 等人，2017年）。

正规化流（NF）的另一个主要应用是变分推断（VI）。考虑参数或潜变量 $l(\boldsymbol{z};\boldsymbol{x})$ 在给定观测数据 $\boldsymbol{z}$ 后，对于 $\boldsymbol{x}_i$， $i=1$, ..., n。给定先验分布 $p(\boldsymbol{z})$，通过最大化对数边际似然 log $q_K(\boldsymbol{z})$ 的下界（即证据下界（ELBO）），或者等价地，通过最小化自由能界（Rezende 和 Mohamed，2015年）可以计算基于正规化流（NF）的后验分布近似 $p(\boldsymbol{z}|\boldsymbol{x})$，其表达式如下：

$$\begin{aligned}\mathcal{F}(\boldsymbol{x}) &= \mathbb{E}_{q_K(\boldsymbol{z})}\left[\log q_K(\boldsymbol{z}) - \log p(\boldsymbol{x},\boldsymbol{z})\right] = \mathbb{E}_{q_K(\boldsymbol{z}_K)}\left[\log q_K(\boldsymbol{z}_K) - \log p(\boldsymbol{x},\boldsymbol{z}_K)\right]\\ &= \mathbb{E}_{q_0(\boldsymbol{z}_0)}[\log q_0(\boldsymbol{z}_0)] - \mathbb{E}_{q_0(\boldsymbol{z}_0)}[\log p(\boldsymbol{x},\boldsymbol{z}_K)] - \mathbb{E}_{q_0(\boldsymbol{z}_0)}\left[\sum_{k=1}^{K}\log\left|\det\frac{\partial F_k}{\partial \boldsymbol{z}_{k-1}}\right|\right]\\ &= \sum_{i=1}^{n}\left(\frac{1}{n}\mathbb{E}_{q_0(\boldsymbol{z}_0)}[\log q_0(\boldsymbol{z}_0)] - \mathbb{E}_{q_0(\boldsymbol{z}_0)}[\log p(\boldsymbol{x}_i,\boldsymbol{z}_K)] - \frac{1}{n}\mathbb{E}_{q_0(\boldsymbol{z}_0)}\left[\sum_{k=1}^{K}\log\left|\det\frac{\partial F_k}{\partial \boldsymbol{z}_{k-1}}\right|\right]\right)\\ &= \textstyle\sum_{i=1}^{n} l_i\end{aligned} \tag{4}$$

方程式（4）中的期望值可通过实现过程中从基础分布 q_0 中抽取的样本 $\boldsymbol{z}_{i,\ 0}$， $i=1$, ..., n 所对应的蒙特卡罗（MC）估计值来近似。

2.2 差分隐私

差分隐私（DP）是一种旨在提供稳健隐私保证的数学框架。它着眼于在最坏情况下保护数据集中每个个体的隐私。

定义 *1*（*(ϵ,δ)-差分隐私*）。（*Dwork* 等人，*2006a,b,c*）若对于任意 $S\in$ 范围*($\mathcal{M}$)* 以及仅相差一条记录的相邻数据集 D_1, D_2，均有 *Pr*$[\mathcal{M}(D_1)\in S] \le e^{\epsilon}$*Pr*$[\mathcal{M}(D_2)\in S]+\delta$，则随机化机制 $\mathcal{M}$ 是 ϵ-差分隐私，其中 $\epsilon>0$ 和 $\delta\in[0,1)$。

通俗地说，如果发布的信息在数据集中包含或不包含某个个体时都大致相同，那么差分隐私意味着，通过随机化机制识别出数据集中某个个体的可能性是有限的

的情况下。量 ϵ 和 δ 表示隐私预算或隐私损失参数。当 $\delta = 0$ 时，定义 ϵ中的δ-差分隐私会退化为纯 1 差分隐私 ϵ。数值 ϵ越小，对数据中个体的隐私保护就越强。

当我们反复应用差分隐私机制来查询数据集时，隐私损失会在这一过程中不断累积。因此，有必要对累积损失进行跟踪和核算，以避免隐私损失过大。这促使人们发展出原始数据差分隐私概念的各种变体和扩展，旨在为组合隐私损失实现更严格的边界，例如集中差分隐私（Bun and Steinke，2016年；Bun et al.，2018年；Dwork andRothblum，2016年）、R´ényi 差分隐私（Mironov，2017年）以及高斯差分隐私（Dong et al.，2019年）。Abadi et al.（2016年）提出了一种基于 R´ényi 差分隐私的、广泛应用的矩会计技术，用于跟踪基于随机梯度下降的优化过程中的隐私损失。Dong etal.（2019年）基于高斯差分隐私（GDP）给出了更严格的边界；高斯差分隐私正是本文后续将采用的差分隐私范式。

定义 *2*（μ-差分隐私）。*(Dong* 等人，*2019)* 设 T 为一个权衡函数，将 $[0,1]$映射到 $[0,1],$，其定义为 $T(P_0, P_1)(\alpha) = inf\{\beta_\phi$：对于来自 P_0的两个概率分布 P_1，通过假设检验（H_0： P_0 与 H_1： P_1）进行区分，其中 $\phi \in [0,1]$ 为拒绝规则，第一类错误为 $\alpha_\phi = \mathbb{E}_{P_0}[\phi]$，第二类错误为 $\beta_\phi = 1 - \mathbb{E}_{P_1}[\phi]$。若随机化机制 $\mathcal{M}$ 满足

$$T(\mathcal{M}(D_1), \mathcal{M}(D_2)) \geq T(\mathcal{N}(0,1), \mathcal{N}(\mu,1))$$

对于任意仅相差一条记录的相邻数据集 D_1、D_2，

在从数据集发布统计量或查询结果时，可以应用高斯机制来实现 μ-差分隐私保证。

定义 3（μ-差分隐私的高斯机制）。（*Dong* 等人，*2019*年）对于给定的查询f 和数据集 D，满足 $\mathcal{M}$（D）的高斯机制 μ-差分隐私定义为

$$f^*_{\mathcal{M}}(D) \sim \mathcal{N}(f(D), \Delta_f^2/\mu^2),$$

其中，对于任意一对相邻数据集最大值$_{D_1,\ D_2}\|f(D_1) - f(D_2)\|_1$。

Dong 等人(2019年) 还提出了一种噪声随机梯度下降（SGD）算法，该算法由重复应用高斯机制并结合子采样组成。如果将 SGD 算法的迭代次数记为 T，子采样率记为 r，每次迭代时添加到梯度上的噪声尺度记为 σ，那么在 GDP 下，经过 T次迭代后的总隐私损失为 $\mu = r\sqrt{T(e^{1/\sigma^2} - 1)}$。

（ϵ，δ）-差分隐私与 μ-GDP 之间存在对偶关系（Dong 等人，2019年），这有助于在这两种差分隐私框架之间转换隐私损失。如果某个机制满足 μ-GDP，那么它也满足（ϵ，$\delta(\epsilon)$)-差分隐私，其中 $\delta(\epsilon) = \Phi(-\frac{\epsilon}{\mu} + \frac{\mu}{2}) - e^{\epsilon}\Phi(-\frac{\epsilon}{\mu} - \frac{\mu}{2}))$，其中 Φ 是标准正态分布的累积分布函数。

3 通过 DP-NF 实现隐私保护密度估计与合成数据生成

本节介绍通过正规化流实现的差分隐私密度估计，可用于在代理数据集公开发布前合成隐私保护的代理数据集。

随后，用户可以像使用原始数据一样，利用合成数据进行统计分析或训练机器学习算法。下文首先介绍启发本研究的电子健康记录数据，然后给出一种将差分隐私融入正规化流以进行密度估计的算法，并应用该算法生成隐私保护的电子健康记录数据，最后通过分类、回归和变分推断等下游学习任务来评估代理数据的效用。

3.1 电子健康记录数据集描述

该电子健康记录数据集包含 82 名成年患者的匿名临床测量数据。这些数据是在一项研究项目中收集的，该项目由 Google 通过其 ATAP 倡议资助，重点是对心血管动力学的无创测量进行建模。表 1 列出了数据集中包含的 26 项临床属性及其汇总统计量。图 1 显示了缺失数据模式。热图中的每一行都归一化到 0 到 1 的范围，以突出显示各临床目标的相对大小。每位患者有 0 至 19 项临床属性。虽然没有任何一位患者具备全部 26 项测量数据，但除一位患者外，其他患者至少都有三项临床属性（心率、舒张压和收缩压）可用。

表 1: 包含 26 项临床测量指标的电子健康记录数据集的汇总统计量

	属性	描述	Unit	SD*	n#	Min	Max
1	心率2	心率	bpm	3.0	81	69.0	105.0
2	收缩压 2	收缩压	mmHg	1.5	81	83.0	180.0
3	舒张压 2	舒张压	mmHg	1.5	81	49.0	102.0
4	心输出量	心输出量	升/分	0.2	65	3.3	10.9
5	全身血管阻力	体循环血管阻力	达因·秒·厘米$^{-5}$	50.0	64	668.6	2046.8
6	肺血管阻力	肺血管阻力	达因·秒·厘米$^{-5}$	5.0	50	5.0	245.8
7	cvp	中心静脉压	mmHg	0.5	30	2.0	28.0
8	右心室舒张期	右心室舒张压	mmHg	1.0	11	2.0	10.0
9	右心室收缩期	右心室收缩压	mmHg	1.0	46	12.0	45.0
10	右心室舒张末期压	右心室舒张末期压	mmHg	1.0	46	1.0	20.0
11	主动脉瓣平均压力梯度	主动脉瓣平均压差	mmHg	0.5	2	5.0	10.0
12	主动脉瓣峰值压力梯度	主动脉瓣峰值压差	mmHg	0.5	37	3.0	18.0
13	二尖瓣减速时间	二尖瓣减速时间	ms	6.0	41	88.0	289.0
14	二尖瓣E/A比值	二尖瓣 E/A 比值	-	0.2	39	0.7	20.9
15	肺动脉加速时间	肺动脉瓣加速时间	ms	6.0	18	53.0	190.0
16	肺动脉最大压力梯度	肺动脉瓣峰值压差	mmHg	0.5	31	1.0	18.0
17	右心房压力	平均右心房压力	mmHg	0.5	50	3.0	15.0
18	右心房四腔心切面容积	右心房容积	mL	3.0	4	57.0	173.0
19	四腔心尖切面左心房容积	左心房容积	mL	3.0	7	69.0	179.0
20	左心室收缩末期容积	左心室收缩末期容积	mL	10.0	1	32.0	32.0
21	四腔心尖切面左心室容积	左心室容积	mL	20.0	5	59.0	642.0
22	lvef	左心室射血分数	-	2.0	53	12.5	75.0
23	左心室流出道最大流量	左心室流出道峰值流速	cm/s	-	0	-	-
24	肺动脉舒张压	肺动脉舒张压	mmHg	1.0	65	5.0	36.0
25	肺动脉收缩压	肺动脉收缩压	mmHg	1.0	65	10.0	60.0
26	楔压	肺动脉楔压	mmHg	1.0	50	2.0	26.0

*标准差 偏差；根据初步文献综述确定（Gordon 等，1983年；Maceira 等，2006；Yared 等，2011年）
#数量 在相应属性上具有观测值的患者。

该数据集聚焦于 II 组肺动脉高压病例（即高血压/无高血压），这类病例中，左心室充盈压升高后，肺血管阻力会出现可逆性升高[1]。实践中常用的一种二元分类方法是基于

[1]我们不考虑肺动脉高压（I 组肺动脉高压），也不声称能够区分 I 组和 II 组肺动脉高压。已有其他标准用于对由左心室舒张功能障碍引起的高血压进行分类（例如参见 Simonneau 等，2013）。

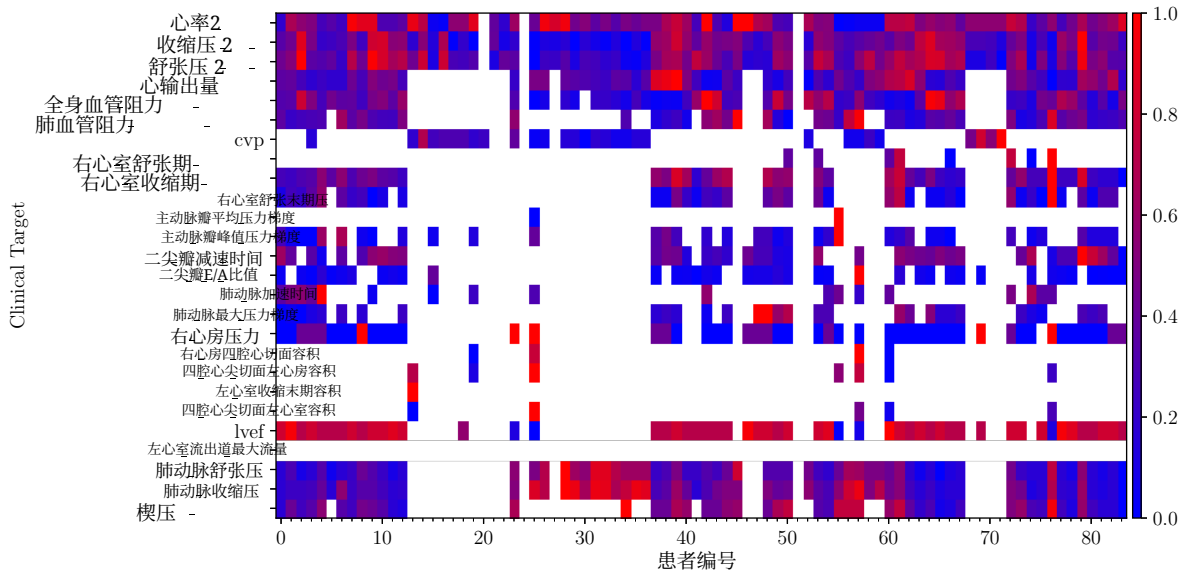


图1：电子健康记录数据集中缺失数据的模式。上述热图中的每一行都归一化到 $[0, 1]$ 范围，以突出显示临床目标在各患者之间的相对大小。

两项临床测量指标：肺动脉舒张压（$p_{p,\ d}$）和收缩期动脉压（$p_{p,\ s}$），即

$$\text{hypertensive} = \begin{cases} 1, & \text{if} \quad p_{p,m} = \frac{2}{3} \cdot p_{p,d} + \frac{1}{3} \cdot p_{p,s} > 20 \text{ mmHg} \\ 0, & \text{if} \quad p_{p,m} = \frac{2}{3} \cdot p_{p,d} + \frac{1}{3} \cdot p_{p,s} \leq 20 \text{ mmHg}. \end{cases} \tag{5}$$

医疗信息（包括电子健康记录数据）高度敏感，如果未妥善保护隐私，可能会泄露患者的医疗状况、病史等信息。由于存在隐私问题，共享这些数据并不可行。研究人员和实践者提出并积极探索生成和发布合成数据，作为一种替代方案（Benaim 等，2020年；Dash 等，2019年；Rankin 等，2020年；SHARED 团队，2023；Yale 等，2020年）。

3.2 通过差分隐私正规化流进行隐私保护合成数据生成

图 1显示，电子健康记录数据集中存在相当比例的缺失值。在生成合成数据之前，我们首先使用 R 中的 `MICE`软件包（v3.13.0）插补缺失值。`MICE` 软件包实现了链式方程多重插补（Van Buuren 和 Groothuis-Oudshoorn，2011年），该方法以交替方式（类似 Gibbs 采样器）从条件分布 $p(X_{\cdot,\ j}|\boldsymbol{\theta}_j,\ \boldsymbol{X}_{\cdot,\ -j})$中抽取缺失属性 $X_{\cdot的j}$ 插补值，并从其后验分布$p(\boldsymbol{\theta}_j|\boldsymbol{X})$中抽取参数 $\boldsymbol{\theta}_j$，适用于 $j = 1, \ldots\ldots, p$。$\boldsymbol{X}_{\cdot,\ -j}$包含数据中的所有变量，但不包括 $X_{\cdot,\ j}$。

在电子健康记录数据中，七项属性（中心静脉压、主动脉瓣平均压力梯度、右心房四腔心切面容积、四腔心尖切面左心房容积、左心室收缩末期容积、四腔心尖切面左心室容积和左心室流出道最大流量）的缺失值比例很高，介于 63% 至 100% 之间。这些高比例的缺失数据对 MICE 来说难以处理。因此，插补时未纳入这七项属性，后续分析也未考虑它们。我们获得了 5 组插补数据，然后将每组

插补数据集划分为训练集和测试集，分别包含 64 名和 20 名患者。

我们使用 DP-NF 在 5 个插补数据集中的每一个上进行密度估计，采用算法 1。在计算出隐私保护密度 $p^*(\boldsymbol{x})$（我们使用 * 来表示隐私或脱敏后的量）后，可以通过以下方式生成合成数据：首先从基分布 $\boldsymbol{x}^*$ 中采样 $\boldsymbol{z}_0 \sim q_0$，然后使用学习到的正规化流映射 $F_{\lambda^*}(\boldsymbol{z}_0)$ 来生成 p^*的实现。尽管该算法采用随机梯度下降框架进行描述，但通过将更新 （Tieleman 等人，2012年）和 Adam（Kingma 和 Ba，2014年）替换为相应的参数更新范式，可以很容易地扩展到其他基于梯度的优化方法，例如 RMSprop $\lambda^{*(t+1)} = \lambda^{*(t)} - \eta\, g^{*(t)}$。

算法1 DP-NF 密度估计

输入：数据集 $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$，NF $F_\lambda(\cdot)$ 的初始参数值 $\lambda^{(0)}$，基分布 q_0，学习率 η，差分隐私噪声尺度 σ，子采样率 r，梯度裁剪常数 C，迭代次数 T。输出：对 $t = 0$、$\cdots$、T 的隐私保护密度估计 $p^*(\boldsymbol{x})$执行子采样：以速率 r从 $\boldsymbol{x}$ 中抽取 $\boldsymbol{x}^{(t)}$；令 $\boldsymbol{x}^{(t)}$的大小为 $b^{(t)}$；对 $\boldsymbol{x}_i \in \boldsymbol{x}^{(t)}$ 执行令 $\boldsymbol{z}_i = F_\lambda^{-1}(\boldsymbol{x}_i)$获取 $l_i = -\log\, q_0(\boldsymbol{z}_i) - \log |\partial F_\lambda / \partial \boldsymbol{z}_i|$计算 $\boldsymbol{g}_i^{(t)} = \nabla_\lambda l_i \boldsymbol{g}_i^{(t)} \leftarrow \boldsymbol{g}_i^{(t)} /$最大值$(1, \|\boldsymbol{g}_i^{(t)}\|_2 / C)$；结束循环；获取脱敏梯度 $\boldsymbol{g}^{*(t)} = \left(\sum_{i=1}^{b^{(t)}} \boldsymbol{g}_i^{(t)} + \mathcal{N}(0, \sigma^2 C^2 \boldsymbol{I})\right) / b^{(t)} \lambda^{*(t+1)} = \lambda^{*(t)} - \eta\, g^{*(t)}$；结束循环

将算法1应用于电子健康记录数据时，我们设置 $r = 0.5, \eta = 2 \times 10^{-5}, T$ =8000, $C = 10$。我们在 μ-GDP 设置下考察了 4 个隐私预算 μ，即 μ =(6.10, 3.92, 2.45, 1.49)[2],，分别对应于 σ =(7.36, 11.44, 18.28, 29.93)；若采用 (ϵ, δ)-DP 设置，则对应于 ϵ =(32, 16, 8, 4)。我们使用的正规化流架构包含 15 个 MADE 层，每层由一个含有 1 个隐藏层、每层有 200 个神经元的全连接神经网络构成，并且每个 MADE 层之后都接一个批归一化层。对于 5 个插补后的训练集，我们分别运行 DP-NF 10 次，以获得 10 个隐私保护密度；从每个估计密度中生成 10 个合成数据集，总计得到 500 个合成数据集[3]。每个合成数据集包含的受试者数量（82）与原始数据集相同。

我们使用 t 生成的合成数据用于三项后续学习和推断任务 s.

[2]我们还尝试了更小的 μ (例如 $\mu \leq 1$)，但由于向梯度中注入了大量噪声，DP-NF算法未能收敛。[3] 我们生成了多个数据集，以考察第 3.3 节和 3.4节中下游学习任务的结果是否稳定。实际应用中，可以发布多份合成数据，以考虑净化和合成过程中的不确定性，然后按照 Liu (2022年) 中的规则合并多份数据得出的推断。相应地，生成隐私保护合成数据时，需要在多份数据之间分配总隐私预算。对于基于合成数据的机器学习任务，发布单个合成数据集可能就足够了。

前两项侧重于检测肺动脉高压，并根据测试数据预测肺动脉压均值（第 3.3 节）；最后一项则对基于物理的模型参数进行变分推断（第 3.4）。

3.3 基于隐私保护合成电子健康记录数据的分类与回归

此学习任务的目标是检测肺动脉高压，其定义为平均肺动脉压力$p_{p,\ m}$在公式（5）中大于 20 毫米汞柱。为此，我们使用支持向量机（SVM，Cortes 和 Vapnik，1995年）训练二元分类器（高血压/非高血压），并使用随机森林（例如，见 Breiman，2001年）估算平均肺动脉压力。我们评估了使用隐私保护合成数据训练的这两种方法在分类和回归测试中的准确率。SVM 会为每个病例输出患肺动脉高压的概率。我们没有采用 0.5 作为阈值，而是使用 Fowlkes–Mallows（FM）指数（Fowlkes 和 Mallows，1983年）来确定最佳阈值。FM 指数定义为 $\sqrt{\text{TP}\ \text{TP+FP}} \cdot \text{TP}\ \text{TP+FN}$，其中 TP 是真正例的数量，FP 是假正例的数量，FN 是假负例的数量。我们采用了能够产生最大 FM 指数的阈值。

表 2 和 3总结了 500 个合成数据集上的预测结果。两张表分别呈现了两组结果：一组基于使用原始生成合成数据训练的 SVM 或随机森林，另一组基于使用后处理合成数据训练的 SVM 或随机森林；后处理方法是将超出边界的合成数据点映射回其下界或上界。进行后处理的原因在于，电子健康记录数据中的属性属于临床属性，具有自然边界。例如，“肺动脉瓣跨瓣峰值压差”这一属性与肺动脉瓣狭窄有关；超出 $[0, 50]$ 的数值将被视为生理上不可能，因为我们的电子健康记录数据集不包含患有肺动脉瓣狭窄的患者。因此，如果“肺动脉瓣峰值压差”的合成值为 < 0 毫米汞柱，则将其设为 0；如果发现为 > 50 毫米汞柱，则将其设为 50 毫米汞柱。

表 2：基于 DP-NF 生成的合成数据，通过 SVM 得到的隐私保护均值分类准确率（标准差）

截断*	插补	原始数据	$\mu = \infty$	$\mu = 6.10$	$\mu = 3.92$	$\mu = 2.45$	$\mu = 1.49$
No	1	1.00	0.95 (0.07)	0.91 (0.08)	0.82 (0.24)	0.81 (0.15)	0.87 (0.09)
	2	1.00	0.88 (0.24)	0.93 (0.06)	0.86 (0.10)	0.88 (0.12)	0.83 (0.11)
	3	1.00	0.92 (0.09)	0.91 (0.06)	0.90 (0.08)	0.88 (0.10)	0.72 (0.38)
	4	0.95	0.89 (0.10)	0.88 (0.05)	0.80 (0.09)	0.76 (0.22)	0.86 (0.12)
	5	0.95	0.85 (0.20)	0.90 (0.07)	0.88 (0.17)	0.92 (0.17)	0.80 (0.10)
	mean	0.98	0.90 (0.16)	0.91 (0.07)	0.85 (0.15)	0.85 (0.20)	0.81 (0.20)
Yes	1	1.00	0.92 (0.08)	0.87 (0.09)	0.79 (0.24)	0.80 (0.15)	0.85 (0.09)
	2	1.00	0.83 (0.26)	0.89 (0.07)	0.86 (0.09)	0.86 (0.11)	0.83 (0.10)
	3	1.00	0.90 (0.11)	0.90 (0.07)	0.90 (0.08)	0.88 (0.10)	0.70 (0.37)
	4	0.95	0.85 (0.16)	0.87 (0.06)	0.80 (0.08)	0.74 (0.33)	0.82 (0.12)
	5	0.95	0.77 (0.28)	0.85 (0.08)	0.85 (0.17)	0.88 (0.17)	0.80 (0.10)
	mean	0.98	0.85 (0.20)	0.88 (0.08)	0.84 (0.15)	0.83 (0.20)	0.80 (0.20)

* 属性的填补数据和合成数据可能超出生理可行界限。“截断 = 是”中的结果是基于后处理的填补数据和合成数据得出的，方法是将超出边界的值设为最接近的边界值。

表 3：基于 DP-NF 生成的合成数据，通过随机森林进行隐私保护预测的均方误差（标准差）

截断*	插补	原始数据	$\mu = \infty$	$\mu = 6.10$	$\mu = 3.92$	$\mu = 2.45$	$\mu = 1.49$
No	1	20.63	40.30 (23.37)	54.71 (20.20)	62.45 (20.55)	83.44 (42.54)	70.39 (30.52)
	2	18.02	35.70 (19.87)	48.08 (18.10)	60.06 (22.89)	62.44 (33.44)	80.30 (44.03)
	3	11.50	71.76 (62.20)	36.53 (7.39)	47.02 (23.01)	72.22 (28.51)	71.36（30.02）
	4	18.87	32.92（27.36）	44.15（12.19）	63.09（19.02）	46.81（26.87）	60.52（17.59）
	5	20.66	81.45 (59.57)	53.98 (17.75)	59.53 (29.41)	41.09 (18.28)	122.91 (41.70)
	mean	17.94	52.52 (47.18)	47.40 (17.17)	58.41 (23.86)	61.39 (34.49)	80.10 (39.92)
Yes	1	21.49	47.50 (27.68)	72.30 (23.98)	79.00 (26.64)	95.76 (43.68)	75.85 (28.33)
	2	18.54	47.08 (24.01)	56.77 (15.78)	68.91 (22.45)	73.43 (36.21)	88.21 (42.44)
	3	12.12	68.25 (55.06)	47.09 (12.74)	55.26 (22.01)	80.08 (29.44)	85.45 (29.94)
	4	18.06	35.55 (30.28)	55.24 (13.12)	71.30 (24.08)	51.26 (26.73)	66.17 (22.38)
	5	19.79	78.79 (53.02)	63.31 (18.31)	68.13 (31.79)	53.75 (21.02)	112.54 (40.61)
	mean	18.00	55.91 (42.61)	58.94 (19.23)	68.52 (26.77)	71.91 (35.70)	84.91 (36.52)

* 属性上的填补数据和合成数据可能超出生理可行界限。“截断 = 是”中的结果是基于后处理后的填补数据和合成数据得出的，方法是将超出边界的值设为最接近的边界值。

总体而言，正如预期，隐私损失越高（ μ越大），准确率（表 2）越高，预测均方误差（表 3）越小。为满足每项临床测量的边界要求而进行的后处理截断符合实际数据约束，但会对预测结果产生负面影响；与不进行截断相比，无论所选隐私预算如何，进行截断都会导致表 2 中的准确率下降，以及表 3 中的预测均方误差上升；但截断后 μ 的趋势与截断前相似。

3.4 使用保护隐私的合成电子健康记录数据对基于物理的模型进行参数变分推断

3.4.1 CVSim-6 生理模型

本研究使用 CVSim-6（Davis，1991；Heldt 等，2010），这是一种可通过 PhysioNet 存储库获取的包含 6 个腔室的集总参数血流动力学模型，用于模拟成年受试者循环系统中的血压、流量和体积的变化（Goldberger 等，2000年）。自定义的 Python/Cython 实现见 https://github.com/desResLab/supplMatHarrod20。该模型包含左心、右心、体循环动脉、体循环静脉、肺动脉和肺静脉腔室。动脉和静脉腔室由 RC 电路构成，而心腔则使用压力发生器、两个单向二极管和一个电阻器进行模拟，以计及瓣膜产生的压力损失。所有元件均视为线性元件，不考虑负压力下静脉的可塌陷性。同时假设惯性效应可以忽略（该模型不包含电感器），并采用双腔心脏模型，不考虑心房的任何作用。此外，仅考虑相邻腔室之间的相互作用。因此，与文献中的其他模型表述不同（例如 CircAdapt，见Arts 等，2005年），不考虑两个心室腔室之间通过室间隔产生的机械相互作用。CVSim-6 由一组包含六个微分方程的系统构成（每个腔室对应一个方程）。在假设瓣膜为非线性单向瓣膜（无反流）的情况下，腔室之间的流量表示为（见Davis，1991，

第4.2.1节和第4.2.2节）

$$q_{li} = \begin{cases} (P_{pv} - P_l)/R_{li} & \text{if } P_{pv} > P_l \\ 0 & \text{otherwise} \end{cases} \qquad q_{lo} = \begin{cases} (P_l - P_a)/R_{lo} & \text{if } P_l > P_a \\ 0 & \text{otherwise} \end{cases}$$

$$q_{ri} = \begin{cases} (P_v - P_r)/R_{ri} & \text{if } P_v > P_r \\ 0 & \text{otherwise} \end{cases} \qquad q_{ro} = \begin{cases} (P_r - P_{pa})/R_{ro} & \text{if } P_r > P_{pa} \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

$$q_a = (P_a - P_v)/R_a \qquad q_{pv} = (P_{pa} - P_{pv})/R_{pv}$$

由此得到电容固定和随时间变化的RC元件微分方程

$$\begin{aligned} \frac{dP_l}{dt} &= \frac{q_{li} - q_{lo} - (P_l - P_{th})\, dC_l(t)/dt}{C_l(t)}, \quad \frac{dP_a}{dt} = \frac{q_{lo} - q_a}{C_a}, \\ \frac{dP_v}{dt} &= \frac{q_a - q_{ri}}{C_v}, \quad \frac{dP_r}{dt} = \frac{q_{ri} - q_{ro} - (P_r - P_{th})\, dC_r(t)/dt}{C_r(t)}, \\ \frac{dP_{pa}}{dt} &= \frac{q_{ro} - q_{pv}}{C_{pa}}, \quad \frac{dP_{pv}}{dt} = \frac{q_{pv} - q_{li}}{C_{pv}}. \end{aligned} \tag{7}$$

初始条件由关于受压体积的线性平衡方程组确定，

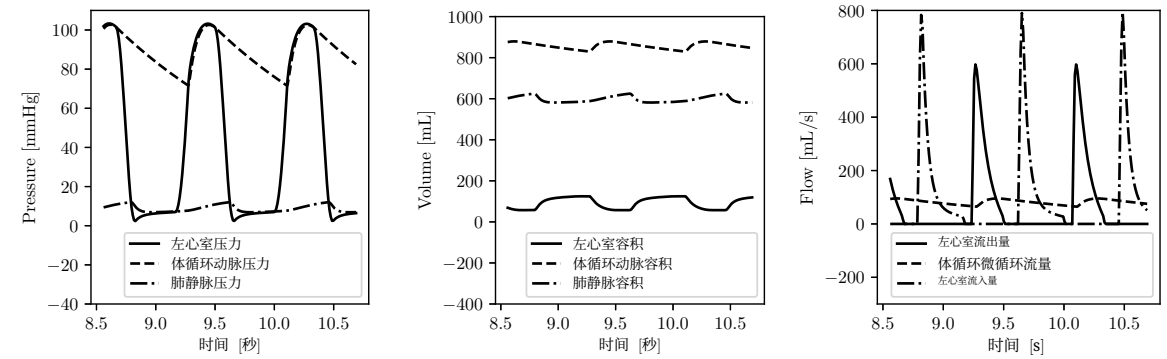
$$\begin{aligned} C_{l,dias}\,(P_{l,dias} - P_{th}) - C_{l,sys}\,(P_{l,sys} - P_{th}) &= C_{r,dias}\,(P_{r,dias} - P_{th}) - C_{r,sys}\,(P_{r,sys} - P_{th}) \\ &= T_{sys}\,\frac{P_{l,sys} - P_a}{R_{lo}} = T_{tot}\,\frac{P_a - P_v}{R_a} \\ &= T_{dias}\,\frac{P_v - P_{r,dias}}{R_v} = T_{sys}\,\frac{P_{r,sys} - P_{pa}}{R_{ro}} \\ &= T_{tot}\,\frac{P_{pa} - P_{pv}}{R_{pv}} = T_{dias}\,\frac{P_{pv} - P_{l,dias}}{R_{li}} \\ V_{tot} - V_{0,tot} &= C_{l,dias}(P_{l,dias} - P_{th}) + C_a(P_a - \frac{1}{3}P_{th}) + \\ &\quad C_v\,P_v + C_r(P_r - P_{th}) + C_{pa}(P_{pa} - P_{th}) + \\ &\quad C_{pv}(P_{pv} - P_{th}). \end{aligned} \tag{8}$$

上述方程中使用的缩略语见表 8 附录。默认模型参数列于表 4 CVSim-6 使用这些参数作为输入时生成的压力、体积和流量时间历程见图 2。

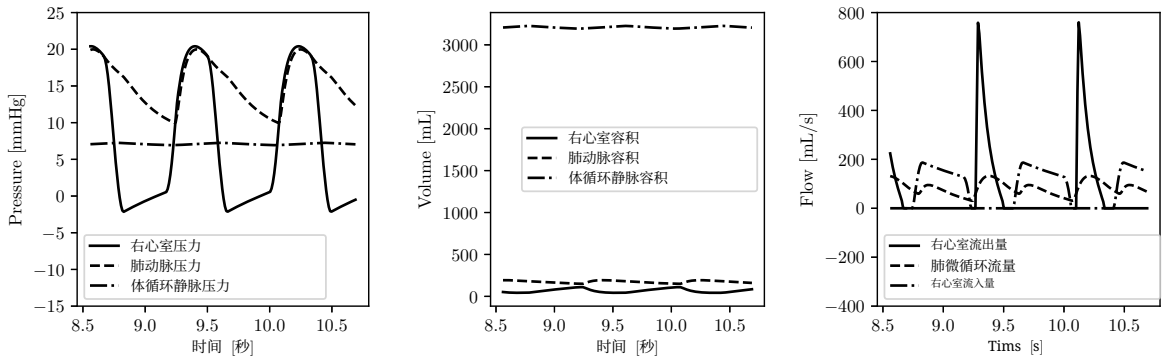
CVSim-6模型可紧凑地表示为 $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{z})$，其中 $\boldsymbol{z}$ 和 $\boldsymbol{x}$ 分别表示模型输入（参数，可以是未知参数或固定常数）和输出（观测值）。在给定电子健康记录数据的情况下，似然函数的构建使用了由8个输出组成的子集，包括心率、肺血管阻力、中心静脉压、右心室舒张压、右心室收缩压、右心室舒张末压、主动脉瓣平均压差和主动脉瓣峰值压差。这些量通过对图2所示的状态变量和辅助变量的时间历程进行后处理获得。在本研究中，我们将右心室阻力 R_{ro} 和动脉电容 C_a 视为未知参数（模型的输入），而其他参数则固定为表4中的默认值。

Ta表4：CVSim-6模型的默认参数（参见 Davis （1991）

腔室	R_i （毫米汞柱·秒/毫升）	R_o （毫米汞柱·秒/毫升）	C（毫升/毫米汞柱）	V_0 （毫升）
左心	0.01	0.006	0.4-10	15
体循环动脉	0.006	1.00	1.6	715
体循环静脉	1.00	0.05	100.0	2500.0
右心	0.05	0.003	1.2-20	15
肺动脉	0.003	0.08	4.3	90
肺静脉	0.08	0.01	8.4	490
体循环	参数：血液总量 = 5000 mL，心率 = 72次/分；胸腔压力 = -4.0 毫米汞柱			.



(a) 左心室压力。 (b) 左心室容积。 (c) 左心室流量。



(d) 右心室压力。 (e) 右心室容积。 (f) 右心室流量。

图2：使用表中列出的默认输入参数时，CVSim-6模型的状态变量和辅助变量的时间历程 4 (参见 Davis，1991)。

3.4.2 基于隐私保护合成电子健康记录数据的变分推断

为了推断高血压患者的CVSim-6模型参数 R_{ro} 和 C_a ，我们基于43名合成高血压患者的子集构建似然函数。我们假设变换参数（R'、C'）在 $[-7,7]^2$上服从均匀先验，

$$(R_{ro},\ C_a) = (\tanh(3/7 \cdot R') \cdot (H_R - L_R)/2 + R_0, \tanh(3/7 \cdot C') \cdot (H_C - L_C)/2 + C_0), \quad (9)$$

其中，$R_{ro} \in [L_R, H_R]$, $C_a \in [L_C, H_C]$ 是具有相应边界的原始参数，R_0，C_0 是默认值，而（R'，C'）是它们经过变换后的无界对应参数。

我们假设，在给定参数 (R_{ro}, C_a) 的情况下，8 维输出 $\boldsymbol{x}_i = (x_{i,\ 1},\ \ldots,\ x_{i,\ 8})$ 服从多元高斯分布。假设这 8 个输出相互独立

wit已知每个输出的边际方差，则 (R_{ro}， C_a) 的对数似然为 s

$$\ell(R_{ro}, C_a; \boldsymbol{x}) = -\frac{1}{2}\sum_{i=1}^{n} \left[\boldsymbol{x}_i - \boldsymbol{f}(R_{ro}, C_a)\right]^T \boldsymbol{\Sigma}^{-1} \left[\boldsymbol{x}_i - \boldsymbol{f}(R_{ro}, C_a)\right]^2 + \text{constant}, \quad (10)$$

其中， $\boldsymbol{f}(R_{ro}, C_a)$表示在除两个输入变量外的所有输入变量均固定为默认值时 CVSim-6 模型的输出，且 $\boldsymbol{\Sigma}$ 是一个对角矩阵，对于 8 维输出，其边际方差为 (444.84, 145.60, 2.21, 45034.41, 475.59, 24.77, 84.06, 15.48)。

CVSim-6 模型的计算开销很大。为降低使用输入参数 R_{ro} 和 C_a进行变分推断的计算开销，我们离线训练了一个全连接神经网络代理模型。该网络包含 2 个隐藏层，分别有 64 和 32 个神经元，并使用双曲正切激活函数。我们在 (R_{ro},C_a) 平面中使用 Sobol' 采样生成了 100 组实现，并计算了对应的 CVSim-6 模型输出。随后，我们基于 100 组输入（2 维）和输出（8 维）数据，使用 RMSprop 优化器进行了 120,000 次迭代训练代理模型（学习率为 0.01，学习率调度器的指数衰减因子为 0.9999），采用 ℓ_2 损失函数和 ℓ_2正则化（惩罚系数为 0.0001）。

我们首先训练代理模型并运行正规化流（NF），以获得变换后的（R'， C'）的后验分布，随后再将其变换回来，以获得 R_{ro}、C_a的后验分布和变分推断（VI）。变分推断采用 MAF架构，通过 5 个交替排列的 MADE 和批归一化层实现。每个 MADE 使用一个全连接神经网络，其特征为包含 100 个神经元的 1 个隐藏层。连续 MADE 层内部及层间的激活函数均为 ReLU。网络训练 15,000 次迭代，每次迭代生成包含 500 个样本的批次。我们使用 RMSProp 作为优化器，学习率设为 0.01，并采用衰减率为 0.9999 的指数调度器。

基于隐私保护合成数据的 R_{ro} 和 C_a 的变分推断结果见表 5，其中展示了两个参数在 5 种不同插补下的后验均值、后验标准差及其相关性。总的来说，基于隐私保护合成数据的估计结果与基于原始数据的估计结果存在一定差异，并且对于μ而言，除了 R_{ro} 和 C_a之间的后验相关性似乎随着 μ（隐私损失）的降低而变得更加负相关之外，前者与后者之间的差异似乎没有明显趋势。

表5：基于 DP-NF 生成的合成数据对 CVSim-6模型参数进行隐私保护变分推断 R_{ro} 和 C_a

参数	后验估计值	原始数据	$\mu = 6.10$	$\mu = 3.92$	$\mu = 2.45$	$\mu = 1.49$
R_{ro}	mean	35.70 (1.918)	43.83 (2.798)	45.90 (2.788)	47.52 (4.400)	45.61 (3.748)
	SD	1.452 (0.038)	1.907 (0.032)	2.168 (0.341)	2.189 (0.383)	2.144 (0.165)
C_a	均值 ($\times 10^{-4}$)	6.520 (0.130)	8.569 (1.169)	8.345 (1.647)	7.972 (0.967)	8.917 (1.023)
	标准差 ($\times 10^{-4}$)	0.347 (0.013)	0.609 (0.135)	0.666 (0.362)	0.594 (0.200)	0.724 (0.192)
R_{ro}， C_a 相关性		-0.120 (0.042)	-0.123 (0.192)	-0.169（0.275）	-0.213（0.244）	-0.193（0.164）

图3 展示了 R_{ro} 和 C_a 的后验样本散点图，这些样本来自在一个私有合成数据集上运行 NF 的结果。观察结果与表5中的结果一致。总体而言，根据原始数据和真实模型构建的正规化流生成的后验样本，来自

隐私保护合成数据生成的后验样本位于基于原始数据和真实模型得到的后验分布高密度区域附近，尽管并未完全聚集在众数周围；对于 $\mu \geq 2.45$，两个参数之间较弱的负相关性也大致得以保留。由于注入噪声以实现差分隐私保证，所有 μ 上的后验样本分布都更加分散，尤其是在 $\mu = 1.49$ 上，这与预期一致。

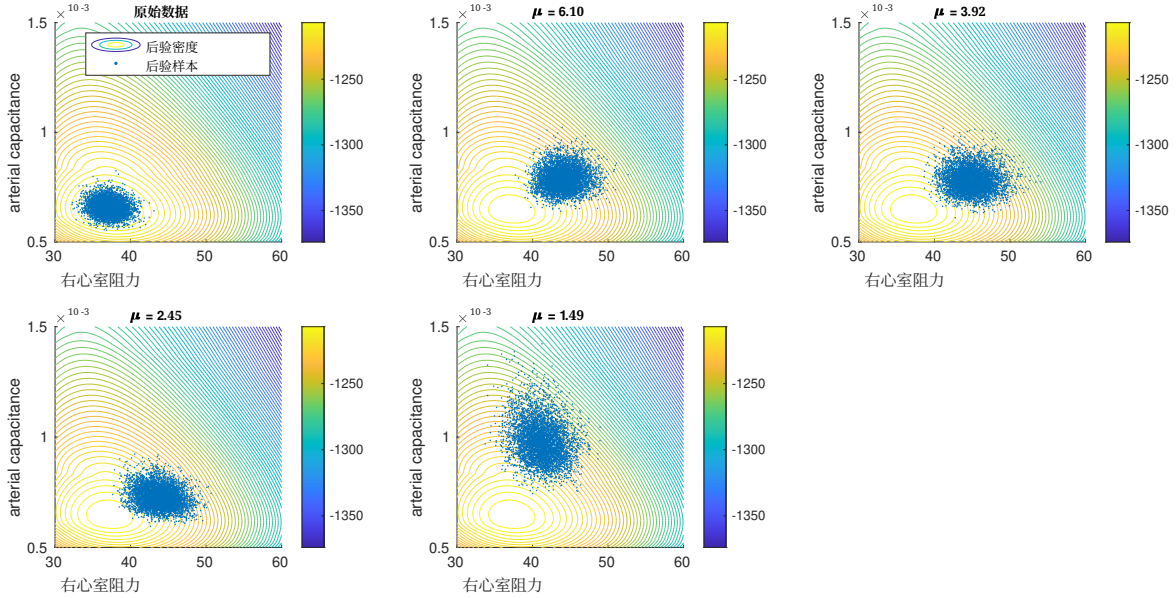


图3：基于一个合成数据集，通过 MAF 得到的 R_{ro} 和 C_a 的隐私保护后验样本示例。等高线表示在真实模型和原始数据条件下， R_{ro} 和 C_a 的真实后验分布密度。

4 通过 DP-NF 实现隐私保护变分推断

本节将探讨通过净化 NF 用于变分推断的优化过程，实现隐私保护变分推断的可行性，简称 DP-NF-VI过程。尽管 NF 可用于密度估计和变分推断，且两种情况下都可通过基于梯度的优化获得相应解，但两者的损失函数不同（式（3）中的似然与式（4）中的自由能界）。下文将首先给出一种 DP-NF-VI 算法，为 NF-VI 提供差分隐私保证；随后，我们将利用模拟数据考察通过 DP-NF 实现隐私保护变分推断的推断效用。

4.1 DP-NF-VI过程

DP-NF-VI过程的步骤如算法2所示。与算法1类似，算法2 通过对梯度进行裁剪并加入加性噪声来扰动梯度，从而由随机梯度下降优化器得到；通过将参数更新步骤（$\lambda^{*(t+1)} = \lambda^{*(t)} - \eta\, g^{*(t)}$）替换为相应的参数更新范式，还可以轻松扩展到其他基于梯度的优化过程，例如 RMSprop 或 Adam。两种算法之间的差异在于损失函数和隐私保护输出：在算法1中，它们分别是自由能和隐私保护变分密度近似 $p^*(\boldsymbol{z})$。我们可以从 $\boldsymbol{z}$ 的隐私保护样本中采样，方法是首先从 $p^*(\boldsymbol{z})$生成 $\boldsymbol{z}_0 \sim q_0$ ，然后应用变换$F_{\lambda^*}(\boldsymbol{z}_0)$。

算法 2 DP-NF-VI过程

输入： 观测数据 $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$，目标后验分布 $p(\boldsymbol{z}|\boldsymbol{x})$，N F $F_\lambda(\boldsymbol{z}) = f_K \circ f_{K-1} \circ \cdots f_1(\boldsymbol{z})$，基分布为 q_0，初始值为 $\lambda^{(0)}$，学习 ng 率为 η，差分隐私噪声尺度为 σ，子采样率为 r，梯度裁剪常数为 C，数量 of 迭代次数为 T，批次大小为 m。输出：隐私保护变分分布 $p^*(\boldsymbol{z})$。对于 $t = 0, \cdots, T$，执行子采样：以 r 的比率从 $\boldsymbol{x}$ 中抽取 $\boldsymbol{x}^{(t)}$；令 $b^{(t)}$ 表示 $\boldsymbol{x}^{(t)}$ 的大小；从 $\boldsymbol{z}_0 \sim q_0$ 中抽取 m 个样本；对于 $\boldsymbol{x}_i \in \boldsymbol{x}^{(t)}$，执行 $l_i(\lambda)$ $= m^{-1} \sum_{j=1}^{m} \{n^{-1} \log\, q_0(\boldsymbol{z}_{0,j}) - \log\, p(\boldsymbol{x}_i, F_\lambda(\boldsymbol{z}_{0,j})) - n^{-1} \log |\partial F_\lambda / \partial \boldsymbol{z}_{0,j}|\}$；得到 $\boldsymbol{g}_i^{(t)} = \nabla_\lambda l_i(\lambda)$ 和 $\boldsymbol{g}_i^{(t)} \leftarrow \boldsymbol{g}_i^{(t)} / \max(1, \|\boldsymbol{g}_i^{(t)}\|_2 / C)$；结束循环；获得脱敏梯度 $\boldsymbol{g}^{*(t)} = \left(\sum_{i=1}^{b^{(t)}} \boldsymbol{g}_i^{(t)} + \mathcal{N}(0, \sigma^2 C^2 \boldsymbol{I})\right) / b^{(t)}$ $\lambda^{*(t+1)} = \lambda^{*(t)} - \eta\, g^{*(t)}$；结束循环

4.2 DP-NF-VI 在模拟数据中的应用

我们使用一个模拟数据集[4] 来评估通过算法 2获得的变分推断的效用。具体而言，我们模拟了输出 $\boldsymbol{x}$，给定协变量 $\boldsymbol{w}$ 和模型参数 $\boldsymbol{\beta}$，这些参数来自以下非线性模型

$$x_{i,k} = f(\boldsymbol{w}_i, \boldsymbol{\beta}) + r_i + e_{i,k}, \tag{11}$$

其中

$$f(\boldsymbol{w}_i, \boldsymbol{\beta}) = \beta_0 + \exp[\beta_1\, w_{i,1}] + \log[w_{i,2} + \exp(\beta_2)] + \exp[\beta_2\, w_{i,3}] + \log[w_{i,4} + \exp(\beta_4)], \tag{12}$$

此外，

$$r_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_{i,k} \sim \mathcal{N}(0, \sigma^2), \tag{13}$$

当 $i = 1, \ldots, 6000$ 且 $k = 1, \ldots 5$ 时。协变量 $\boldsymbol{w}$ 通过 $w_{i,j} \sim \mathcal{U}(0, a_j)$ 独立生成，其中 $i = 1, \cdots, n$ 和 $a_1 = 1$, $a_2 = 3$, $a_3 = 0.5$, $a_4 = 2$。我们设定 $\sigma_0 = \sigma = 0.2$ 以及 $\boldsymbol{\beta} = (0.2, 1.0, 0.8, -1.2, 0.6)^T$。随机效应 r_i 在同一受试者的 5 次重复测量之间引入相关性，其中该受试者的编号为 i，相关性为$\sigma^2/(\sigma^2 + \sigma_0^2) = 1/2$。

我们的目标是在给定非信息均匀先验 $p(\boldsymbol{\beta})$ 和似然 $l(\boldsymbol{\beta}|\boldsymbol{x})$ 的情况下，估计 $\boldsymbol{\beta}$ 的隐私保护后验分布，并获得其隐私保护后验推断；其中，在不失一般性的情况下，假设 σ_0^2、σ^2 已知。我们将算法 2 应用于模拟数据，在不同级别的 μ-差分隐私保证下进行实验（$\mu = 6.58$、1.12、0.5 和 0.27；当 $\delta = 1 \times 10^{-5}$ 时，它们分别对应于 $\epsilon = 50$、5、2 和 1）。我们采用了包含一个 MADE（1 个隐藏层、10 个神经元）和一个批归一化层的 MAF架构。我们还使用大小为 $m = 1,000$ 的批次来计算算法 2 中损失函数的蒙特卡洛估计，子采样率为 $r = 500/6,000$，学习率为 $\eta = 0.01$，并采用指数调度器衰减因子 0.999，迭代次数为 $T = 8,000$，裁剪常数为 $C = 10.0$。

[4]我们还在电子健康记录数据集上运行了该算法，但未能收敛。

关于 $\boldsymbol{\beta}$ 的隐私保护后验推断结果见表 6。总的来说，无论选择何种隐私损失参数 μ，隐私保护后验均值基本保持不变，而括号中的标准差值大于相应的非隐私值（“原始 VI NF”列），并且随着 μ 减小而逐渐增大。此外，使用 DP-NF-VI 计算得到的 $\boldsymbol{\beta}$ 不同元素之间的隐私相关性通常比相应的原始值更明显，并且在 5 次不同的 DP-NF-VI 运行中伴随较大的标准差，表明这些估计总体上缺乏稳定性。

表 6：非线性回归模型参数的差分隐私变分推断 $\boldsymbol{\beta}$ 来自 DP-NF-VI（算法 2）。

参数	原始数据 非线性模型	原始数据 NF-VI	$\mu = 6.68$ $\sigma = 1.30$	$\mu = 1.12$ $\sigma = 6.68$	$\mu = 0.50$ $\sigma = 14.88$	$\mu = 0.27$ $\sigma = 27.82$
后验均值						
β_0 (0.2)	0.1581	0.1593（0.0020）	0.1532 (0.0035)	0.1381（0.0151）	0.1231 (0.0486)	0.1419（0.0701）
β_1 (1.0)	1.0022	1.0016（0.0015）	0.9952 (0.0040)	0.9957（0.0036）	0.9922 (0.0059)	0.9943（0.0026）
β_2 (0.8)	0.8262	0.8265 (0.0013)	0.8386 (0.0034)	0.8404 (0.0150)	0.8378 (0.0208)	0.8584（0.0278）
β_3 (-1.2)	-1.1909	-1.1920 (0.0008)	-1.1928（0.0019）	-1.1891 (0.0079)	-1.1822（0.0201）	-1.1945（0.0246）
β_4 (0.6)	0.6088	0.6092（0.0006）	0.6100（0.0010）	0.6215（0.0057）	0.6311 (0.0390)	0.6057 (0.0479)
后验标准差						
β_0	0.0265	0.0255 (0.0011)	0.1341 (0.0147)	0.2576（0.0241）	0.2573（0.0490）	0.2791（0.0825）
β_1	0.0036	0.0036（0.0001）	0.1089（0.0884）	0.0687（0.0587）	0.1344（0.1247）	0.0554（0.0058）
β_2	0.0179	0.0174（0.0008）	0.0755 (0.0222)	0.1267 (0.0250)	0.1729（0.0309）	0.1655 (0.0305)
β_3	0.0151	0.0151（0.0004）	0.0812（0.0056）	0.1341（0.0141）	0.1644 (0.0203)	0.1581 (0.0217)
β_4	0.0192	0.0191 (0.0006)	0.0889 (0.0163)	0.1197 (0.0089)	0.1595（0.0384）	0.2012（0.0399）
后验相关性						
β_0, β_1	-0.1200	-0.0319(0.0499)	-0.1147（0.2240）	-0.1230 (0.2623)	0.0237（0.1028）	-0.1619（0.2592）
β_0, β_2	-0.4340	-0.3415（0.1899）	-0.1180（0.5120）	-0.3923（0.4756）	0.0396（0.3675）	-0.3269（0.7675）
β_0, β_3	-0.7370	-0.7550(0.0127)	-0.9108（0.0851）	-0.8923（0.0353）	-0.8629（0.0407）	-0.8622 (0.0420)
β_0, β_4	-0.4880	-0.4883（0.0356）	-0.5299 (0.4673)	-0.8545（0.0839）	-0.6826 (0.1586)	-0.4920 (0.5249)
β_1, β_2	-0.0030	-0.0082（0.0417）	-0.0310（0.5503）	0.0285（0.0406）	0.0916（0.1965）	-0.0395（0.1101）
β_1, β_3	-0.0160	-0.0892(0.0264)	0.1255 (0.2115)	-0.0381 (0.2270)	-0.1815（0.1485）	-0.0691 (0.2162)
β_1, β_4	-0.0080	-0.0680(0.0252)	-0.2880 (0.4993)	0.1445（0.2869）	-0.0457（0.0595）	0.0784 (0.2416)
β_2, β_3	0.0040	-0.0538(0.1526)	-0.0947 (0.6286)	0.1131 (0.4442)	-0.2520 (0.4011)	0.1453（0.7946）
β_2, β_4	-0.0020	-0.0836(0.0936)	-0.2267（0.4445）	0.3336（0.5772）	-0.2036（0.2303）	-0.0271 (0.5412)
β_3, β_4	-0.0020	0.0318 (0.0443)	0.3368 (0.3616)	0.6469 (0.1602)	0.4218 (0.0884)	0.2300（0.5860）
结果为 DP-NF-VI 运行 5 次的平均值；括号内为这 5 次运行的标准差值						.

作为对比，我们还对 $\boldsymbol{\beta}$ 执行了变分推断，所用的非线性模型基于使用 DP-NF 进行密度估计所生成的隐私保护合成数据（算法 1）。对于此任务，我们使用由一个 MADE（1 个隐藏层，10 个神经元）和一个批归一化层组成的 MAF架构。我们还使用了 RMSprop，批次大小为 500，学习率为 0.01，并采用衰减因子为 0.999 的指数学习率调度器。迭代总次数设为 8,000，梯度裁剪常数 C 为 10.0。

结果见表7。总体而言，无论 μ如何，基于隐私保护合成数据的 NF-VI 所得到的$\boldsymbol{\beta}$ 后验估计都表现出更高的准确性，并且比直接通过 DP-NF-VI 获得的结果更接近非隐私推断的结果（表6）。此外，多组合成数据中括号内的标准差值相对较小，而后验相关性 $\boldsymbol{\beta}$ 与其原始的非隐私值相当接近。

表7：基于通过 DP-NF 生成的合成数据进行的隐私保护变分推断，用于密度估计（算法 1）。

参数	原始数据 非线性 模型	原始数据 NF	$\mu = 6.68$ $\sigma = 1.30$	$\mu = 1.12$ $\sigma = 6.68$	$\mu = 0.50$ $\sigma = 14.88$	$\mu = 0.27$ $\sigma = 27.82$
后验均值						
β_0 (0.2)	0.1581	0.1593（0.0020）	0.1632（0.0146）	0.1062 (0.0505)	0.0866（0.0127）	0.0448（0.0600）
β_1 (1.0)	1.0022	1.0016 (0.0015)	0.9891 (0.0060)	0.9943（0.0060）	0.9917（0.0097）	0.9775 (0.0126)
β_2 (0.8)	0.8262	0.8265 (0.0013)	0.8235 (0.0078)	0.8357（0.0286）	0.8246 (0.0230)	0.8294 (0.0515)
β_3 (-1.2)	-1.1909	-1.1920 (0.0008)	-1.1807 (0.0081)	-1.1630（0.0251）	-1.1406 (0.0138)	-1.1034 (0.0225)
β_4 (0.6)	0.6088	0.6092（0.0006）	0.6011（0.0230）	0.6261 (0.0237)	0.6197 (0.0145)	0.6214（0.0233）
后验标准差						
β_0	0.0265	0.0255 (0.0011)	0.0280 (0.0004)	0.0279（0.0003）	0.0277（0.0006）	0.0274 (0.0004)
β_1	0.0036	0.0036（0.0001）	0.0040 (0.0000)	0.0040 (0.0000)	0.0039（0.0001）	0.0038 (0.0001)
β_2	0.0179	0.0174（0.0008）	0.0201（0.0003）	0.0210 (0.0005)	0.0207（0.0013）	0.0203（0.0017）
β_3	0.0151	0.0151（0.0004）	0.0159（0.0003）	0.0153 (0.0002)	0.0144（0.0007）	0.0142（0.0006）
β_4	0.0192	0.0191 (0.0006)	0.0208 (0.0002)	0.0229（0.0003）	0.0247 (0.0020)	0.0250（0.0020）
后验相关性						
β_0, β_1	-0.1200	-0.0319（0.0499）	-0.0590 (0.0262)	-0.0568（0.0130）	-0.0647 (0.0338)	-0.0359 (0.0185)
β_0, β_2	-0.4340	-0.3415（0.1899）	-0.4184（0.0159）	-0.4155（0.0378）	-0.4588 (0.0606)	-0.4774（0.0296）
β_0, β_3	-0.7370	-0.7550（0.0127）	-0.7293 (0.0050)	-0.6919（0.0092）	-0.6951（0.0200）	-0.7049（0.0326）
β_0, β_4	-0.4880	-0.4883 (0.0356)	-0.5206（0.0103）	-0.5418（0.0122）	-0.5168 (0.0315)	-0.4835 (0.0388)
β_1, β_2	-0.0030	-0.0082 (0.0417)	-0.0289（0.0211）	-0.0244（0.0196）	-0.0315（0.0297）	-0.0774（0.0207）
β_1, β_3	-0.0160	-0.0892（0.0264）	-0.0723 (0.0285)	-0.0674 (0.0176)	-0.0607 (0.0249)	-0.0557 (0.0267)
β_1, β_4	-0.0080	-0.0680 (0.0252)	-0.0269 (0.0204)	-0.0318 (0.0166)	-0.0229 (0.0198)	-0.0428 (0.0171)
β_2, β_3	0.0040	-0.0538（0.1526）	-0.0202（0.0194）	-0.0409（0.0257）	0.1278 (0.2343)	0.2144 (0.2168)
β_2, β_4	-0.0020	-0.0836 (0.0936)	-0.0346（0.0156）	-0.0518 (0.0347)	-0.1289（0.0761）	-0.1792（0.0979）
β_3, β_4	-0.0020	0.0318 (0.0443)	0.0484（0.0110）	0.0316（0.0170）	-0.0257（0.0677）	-0.0676（0.0510）

结果是对通过 DP-NF 进行密度估计所生成的 5 组合成数据取平均得到的；标准差值
括号内的数值是基于这 5 组数据计算得到的。

4.3 DP-NF-VI 的进一步效用分析

表6中所示的参数间后验相关性结果并不理想，这并非偶然。我们在不同数据集上进行了多次实验，以测试 DP-NF-VI算法，并观察到了类似的结果。我们推测，这可能是由于在 DP-NF-VI 中使用自由能界作为损失函数所致。在 NF-VI 中，最小化自由能界等价于最小化真实后验与其变分近似之间的 KL散度。由于在实现差分隐私保证的差分隐私梯度优化中注入噪声并进行梯度裁剪，相较于非隐私目标分布，隐私保护变分分布中的参数后验方差预计会增大。因此，DP-NF-VI过程会在变分族中寻找通过 KL散度衡量与真实后验最相似、但具有更大方差的密度。这可能导致最优变分密度具有不同于原始分布的相关结构。

为进一步展示这一现象并检验我们的猜想，我们进行了以下实验。尽管实验很简单，但结果颇具趣味，也富有启发性。假设目标分布是二元高斯分布

$$p = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_0), \text{ with } \mathbf{\Sigma}_0 = \sigma_0^2 \left[(1-\rho_0)\boldsymbol{I}_{2\times 2} + \rho_0 \mathbf{1}_{2\times 2}\right].$$

We 旨在用另一个二元高斯分布来近似该分布 n

$$q = \mathcal{N}(\mathbf{0}, \Sigma), \text{ with } \boldsymbol{\Sigma} = \sigma^2 \left[(1-\rho)\, \boldsymbol{I}_{2\times 2} + \rho \mathbf{1}_{2\times 2}\right],$$

可被视为隐私分布；与 p 中的分布相比，预计其在每个维度上的边际方差都更大；即 $\sigma > \sigma_0$。我们考察 p 与 q之间的以下四种差异度（或散度）度量

$$L(p,q) = \begin{cases} \text{KL divergence} & D(q\|p), \\ \text{reversed KL divergence} & D(p\|q), \\ \ell_1 \text{ distance}^{5} & \|q-p\|_1 = \int_{\mathbb{R}^2} |p(\boldsymbol{x}) - q(\boldsymbol{x})|\, d\boldsymbol{x}, \\ \text{Wasserstein distance} & W_\alpha(p,q) = (\inf_{\gamma\in\Gamma(p,q)} \mathbb{E}_\gamma |p-q|^\alpha)^{1/\alpha}. \end{cases}.$$

在本实验中，Wasserstein 距离计算如下

$$W_2(p,q) = \text{trace}(\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma} - 2[\boldsymbol{\Sigma}^{1/2}\, \boldsymbol{\Sigma}_0\, \boldsymbol{\Sigma}^{1/2}]^{1/2}).$$

我们计算了 p 与 q 之间的差异度（损失函数），将上述 4 种度量分别表示为 ρ 和 σ 的函数，并在图 4中绘制了相应的等高线图，分别针对三对（ρ_0，σ_0）取值为（0，1）、（1/2，1）和（1/2，4）。随后，在给定膨胀后的标准差 $\sigma = \hat{\sigma}$的情况下，我们确定了使差异度最小的 ρ 的取值；即，

$$\rho^*(\hat{\sigma}) = \arg\min_\rho L(\rho_0, \sigma_0, \rho, \hat{\sigma}),$$

图 4 中的实验结果表明，具有 ρ 相关性的隐私分布，其膨胀后的方差（$\sigma > \sigma_0$）在差异度最小的意义下最优地逼近目标分布时，通常取决于 ρ_0 以及所选的差异度量。

当 $\rho_0 = 0$ 时，通过 KL 散度和 Wasserstein 距离确定的最优隐私近似都能正确匹配目标相关性，即 $\rho = 0$；若使用反向 KL 散度或 ℓ_1 距离，最优隐私近似甚至不唯一，除非其边际方差 σ 足够接近或小于 σ_0。

当 ρ_0 不为 0 时，具有膨胀方差的最优隐私近似表现为相关性 ρ，并且随着膨胀比率 σ/σ_0 增大而迅速趋近于 1.0。在 KL 散度、反向 KL 散度和 ℓ_1 距离的情况下，均可观察到这一现象。然而，无论膨胀方差 σ 如何，Wasserstein 距离仍能得到具有 $\rho = \rho_0$ 的隐私近似。

总之，如果目标是在私有变分概率分布中保留估计参数之间的相关结构，那么 Wasserstein 距离似乎是最佳度量，即便为了实现隐私保证而进行噪声注入会导致边际方差膨胀，也依然如此。

5 讨论

我们研究了通过隐私保护 NF 生成合成数据以进行密度估计的可行性和效用，并将其应用于一个电子健康记录数据集。总体而言，结果

[4]等价于全变差范数的两倍 $\|q-p\|_{TV} = \sup_{A\subset\mathbb{R}^2} |p(A) - q(A)|$

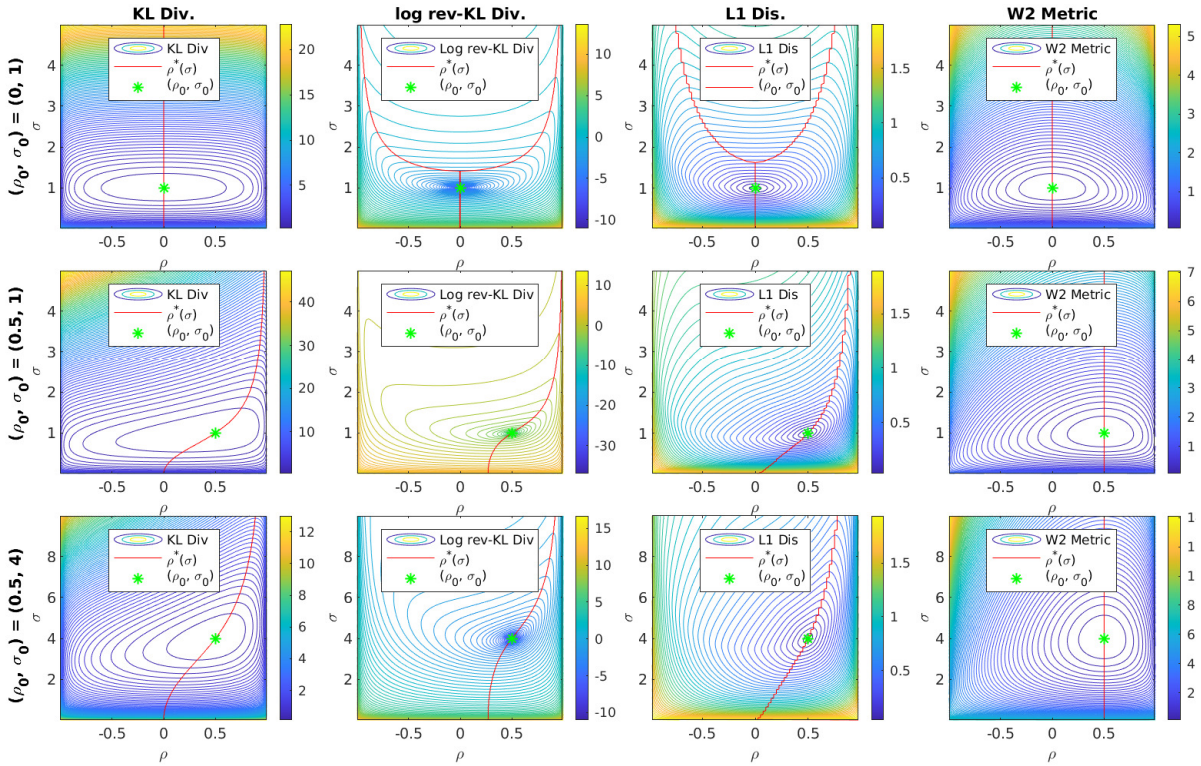


图4：目标分布 $p = \mathcal{N}$(目标0， Σ_0) 与其隐私保护近似分布 $q = \mathcal{N}$(隐私保护0， Σ) 之间的 KL 散度、反向 KL 散度、 ℓ_1 和 Wasserstein距离。等高线图中的红线表示 $\rho^*(\hat{\sigma}) = \arg\min_\rho L(\rho_0, \sigma_0, \rho, \hat{\sigma})$ 的轨迹，该轨迹使给定 $\hat{\sigma}$ 时损失函数 L 最小。

表明，对于该数据集，在实际合理的隐私损失参数取值下，基于隐私保护合成数据进行的下游学习（预测）和推断程序（变分推断）能够取得效用令人满意且可接受的结果。尽管还需要更多实证研究，但这项针对真实数据集的研究发现，该数据集的样本量相对较小，属性数量相对较多，且属性类型混杂并存在缺失值；研究结果表明，通过隐私保护 NF 生成合成数据以进行密度估计，可能是发布和共享敏感信息的一种切实可行的选择。

我们还研究了直接对 NF-VI 程序进行隐私净化，以实现隐私保护后验推断的可行性。我们表明，将自由能界作为 NF-VI 的损失函数，可能会导致变分分布显著偏离目标分布。这意味着，其他损失函数或许能更好地保留目标分布某些方面的信息，例如其相关结构。在此背景下，我们计划进一步探究不同距离度量对 DP-NF-VI 算法性能的影响，以此制定更有效的策略，例如采用基于Wasserstein距离的损失形式。

通过 DP-NF 实现密度估计和变分推断的代码可在 GitHub 上获取，网址为 https://github.com/cedricwangyu/DPNF。

致谢

本研究得到 NSF大数据科学与工程资助 #1918692（首席研究员：DES）的支持，

美国国家科学基金会 CAREER 项目 #1942662（首席研究员：DES）、美国国家科学基金会 OAC CDS&E 项目 #2104831（圣母大学首席研究员：DES）的资助，并使用了圣母大学研究计算中心提供的计算资源。

附录

A. 伦理标准合规性

本研究使用的电子健康记录数据集包含来自涉及人类参与者的外部研究的数据。这些研究中开展的医疗程序均遵循机构和国家研究委员会的伦理标准，以及1964年《赫尔辛基宣言》及其后续修订。本研究被归类为不涉及人类受试者的研究，并于2019年6月13日经圣母大学研究合规办公室和机构审查委员会批准，批准编号为IRB#19-05-5371。

B. CVSim-6模型中的缩略语

表8：CVSim-6模型中的缩略语

流量			
q_{li}	左心室流入	q_{lo}	左心室流出
q_{ri}	右心室流入	q_a	体循环主动脉血流
q_{ro}	右心室流出	q_{pv}	肺静脉血流
压力			
P_{pv}	肺静脉压力	P_l	左心室压
P_a	体循环动脉压力	P_v	体循环静脉压力
P_r	右心室压	P_{pa}	肺动脉压力
P_{th}	胸腔压力	$P_{r,dias}$	右心室舒张压
$P_{r,sys}$	右心室收缩压	$P_{l,dias}$	左心室舒张压
$P_{l,sys}$	左心室收缩压		
阻力			
R_a	体循环动脉阻力	$R_v = R_{ri}$	体循环静脉阻力
R_{lo}	主动脉瓣阻力	R_v	体循环静脉阻力
R_{ro}	肺动脉瓣阻力	$R_{pv} = R_{li}$	肺静脉阻力
顺应性			
$C_l(t)$	左心室顺应性	$C_r(t)$	右心室顺应性
C_a	体循环动脉顺应性	C_v	体循环静脉顺应性
C_{pa}	肺动脉顺应性	C_{pv}	肺静脉顺应性
$C_{l,dias}$	左心室舒张期顺应性	$C_{l,sys}$	左心室收缩期顺应性
$C_{r,dias}$	右心室舒张期顺应性	$C_{r,sys}$	收缩期
时间			
T_{tot}	心动周期总时长	T_{dias}	舒张期时长
T_{sys}	收缩期时长		
容积			
V_{tot}	血液总量	$V_{0,tot}$	非应激血容量

C. 超参数

我们在表中列出了所有实验所用的超参数 9。EHR DP-DE 表示对电子健康记录数据进行差分隐私密度估计（见第 3.2节）；EHR VI 表示在隐私保护合成电子健康记录数据上进行带 KL 散度的变分推断（见第 3.4.2节）；Reg DP-VI

表示通过非线性回归模型对模拟数据进行差分隐私变分推断（见第 4.2节）；Reg DP-DE 表示对模拟数据进行差分隐私密度估计（见第 4.2节）；Reg VI 表示在隐私保护模拟数据上进行带 KL 散度的变分推断（见第 4.2节）。除 CVSim-6 模型的代理模型（全连接神经网络）使用双曲正切作为激活函数外，所有其他正规化流和神经网络均使用 ReLU 作为激活函数。

表9：所有实验中使用的超参数

参数	EHR DP-DE	EHR VI	RegDP-VI 回归差分隐私密度估计		Reg 变分推断
块数	15	5	1	18	1
隐藏单元数	1	1	1	1	1
隐藏层大小	200	100	10	100	10
输入大小	19	2	5	9	5
迭代次数	8000	15000	8000	8000	8000
批次大小	100	500	1000	100	1000
优化器	SGD	RMSProp	RMSProp	RMSProp	RMSProp
学习率	0.00002	0.01	0.01	0.002	0.01
调度器	-	Exp	Exp	Exp	Exp
衰减因子	-	0.9999	0.999	0.9995	0.999
裁剪	10	-	10	5	-
泊松率	50%	-	8.33%	8.33%	-

参考文献

M. Abadi、A. Chu、I. Goodfellow、H. B. McMahan、I. Mironov、K. Talwar 和 L. Zhang。差分隐私深度学习。载于*2016*年 *ACM SIGSAC* 计算机与通信安全会议论文集，第308–318页，2016年。

C. C. Aggarwal 和 P. S. Yu。隐私保护数据挖掘模型与算法综述。载于隐私保护数据挖掘，第11–52页。施普林格，2008年。

F. Alda 和 B. I. Rubinstein。伯恩斯坦机制：差分隐私下的函数发布。载于第三十一届*AAAI*人工智能大会，2017年。

T. Arts、T. Delhaas、P. Bovendeerd、X. Verbeek 和 F. W. Prinzen。对机械负荷的适应决定了心脏和循环系统的形态与特性：CircAdapt模型。美国生理学杂志——心脏与循环生理学，288(4):H1943–H1954，2005年。

M. Backes、P. Berrang、M. Humbert 和 P. Manoharan。基于微小RNA的研究中的成员隐私。载于*2016*年 *ACM SIGSAC* 计算机与通信安全会议论文集，第 319–330 页，2016年。

B. K. Beaulieu-Jones、Z. S. Wu、C. Williams、R. Lee、S. P. Bhavnani、J. B. Byrd 和 C. S. Greene。隐私保护的生成式深度神经网络支持临床数据共享。《循环：心血管质量与结局》，12(7):e005122，2019年。

A. R. Benaim、R. Almog、Y. Gorelik、I. Hochberg、L. Nassar、T. Mashiach、M. Khamaisi、Y. Lurie、Z. S. Azzam、J. Khoury 等。基于合成数据分析医学研究结果及其与真实数据结果的关系：来自五项观察性研究的系统比较。*JMIR* 医学信息学，8(2):e16492，2020年。

C. M. Bowen 和 F. Liu。差分隐私数据合成方法的比较研究。*Statistical Science*，35(2):280–307，2020年。L. Breiman。随机森林。机器学习，45(1):5–32，2001年。Z. Bu、J. Dong、Q. Long 和 W. J. Su。基于高斯差分隐私的深度学习。*Harvard Data Science Review*，2020(23)，2020年。M. Bun 和 T. Steinke。集中差分隐私：简化、扩展与下界。载于 密码学理论会议，第 635–658 页。施普林格，2016年。M. Bun、C. Dwork、G. N. Rothblum 和 T. Steinke。通过截断集中差分隐私实现可组合且灵活的隐私。载于第 *50* 届 *ACM SIGACT* 计算理论研讨会论文集，第 74–86 页，2018年。C. Butucea、A. Dubois、M. Kroll 和 A. Saumard。局部差分隐私：最优密度估计中的肘部效应及 Besov 椭球上的自适应性。*Bernoulli*，26(3):1727–1764，2020年。Q. Chen、C. Xiang、M. Xue、B. Li、N. Borisov、D. Kaarfar 和 H. Zhu。差分隐私数据生成模型。*arXiv* 预印本 *arXiv:1812.02274*，2018年。C. Cortes 和 V. Vapnik。支持向量网络。机器学习，20(3):273–297，1995年。S. Dash、R. Dutta、I. Guyon、A. Pavao、K. P. Bennett 等。保护隐私的合成健康数据。载于*ESANN 2019*——欧洲人工神经网络、计算智能与机器学习研讨会，2019年。T. L. Davis。通过血流动力学交互式模拟教授生理学。博士论文，麻省理工学院，1991年。L. Dinh、J. Sohl-Dickstein 和 S. Bengio。使用 real NVP 进行密度估计。*arXiv* 预印本 *arXiv:1605.08803*，2016年。J. Dong、A. Roth 和 W. J. Su。高斯差分隐私。*arXiv* 预印本 *arXiv:1905.02383*，2019年。C. Dwork 和 G. N. Rothblum。集中差分隐私。*arXiv:1603.01887v2*，2016年。C. Dwork、K. Kenthapadi、F. McSherry、I. Mironov 和 M. Naor。我们的数据，我们自己：通过分布式噪声生成实现隐私。载于密码学进展：*EUROCRYPT* 论文集，第 485–503 页。Springer Berlin Heidelberg，2006a。C. Dwork、K. Kenthapadi、F. McSherry、I. Mironov 和 M. Naor。我们的数据，我们自己：通过分布式噪声生成实现隐私。载于 密码技术理论与应用国际年会，第 486–503 页。施普林格，2006b。C. Dwork、F. McSherry、K. Nissim 和 A. Smith。在隐私数据分析中根据敏感度校准噪声。载于密码学理论会议，第 265–284 页。施普林格，2006c。C. Dwork、A. Smith、T. Steinke 和 J. Ullman。曝光！针对隐私数据的攻击综述。*Annu. Rev. Stat. Appl*，4(1):61–84，2017年。E. B. Fowlkes 和 C. L. Mallows。比较两个层次聚类的方法。《美国统计学会杂志》，78(383):553–569，1983年。

M. Germain、K. Gregor、I. Murray 和 H. Larochelle。Made：用于分布估计的掩码自编码器。载于《国际机器学习会议》，第 881–889 页。PMLR，2015年。A. L. Goldberger、L. A. Amaral、L. Glass、J. M. Hausdorff、P. C. Ivanov、R. G. Mark、J. E. Mietus、G. B. Moody、C.-K. Peng 和 H. E. Stanley。PhysioBank、PhysioToolkit 和 PhysioNet：复杂生理信号新研究资源的组成部分。《循环》，101(23):e215–e220，2000年。I. Goodfellow、J. Pouget-Abadie、M. Mirza、B. Xu、D. Warde-Farley、S. Ozair、A. Courville 和 Y. Bengio。生成对抗网络。载于《神经信息处理系统进展》，第 27 卷，2014年。E. P. Gordon、I. Schnittger、P. J. Fitzgerald、P. Williams 和 R. L. Popp。二维超声心动图测量左心室容积的可重复性。《美国心脏病学会杂志》，2(3):506–513，1983年。R. Hall、A. Rinaldo 和 L. Wasserman。函数与函数型数据的差分隐私。《机器学习研究杂志》，14(1):703–727，2013年。K. K. Harrod、J. L. Rogers、J. A. Feinstein、A. L. Marsden 和 D. E. Schiavazzi。左心室舒张功能障碍继发性肺动脉高压的预测建模。《生理学前沿》，12，2021。网址：https://www.frontiersin.org/article/10.3389/fphys.2021.666915。T. Heldt、R. Mukkamala、G. B. Moody 和 R. G. Mark。CVSim：用于教学和研究的开源心血管模拟器。《开放式起搏、电生理与治疗杂志》，3:45，2010年。N. Homer、S. Szelinger、M. Redmann、D. Duggan、W. Tembe、J. Muehling、J. Pearson、D. Stephan、S. Nelson 和 D. Craig。利用高密度SNP基因分型微阵列，从高度复杂混合物中识别贡献微量DNA的个体。《PLoS Genetics》，4:e1000167，2008年。S. Ioffe 和 C. Szegedy。批归一化：通过减少内部协变量偏移加速深度网络训练。载于《国际机器学习会议》，第 448–456 页。PMLR，2015年。J. J¨alk¨o、O. Dikmen 和 A. Honkela。非共轭模型的差分隐私变分推断。arXiv 预印本 arXiv:1610.08749，2016年。J. J¨alk¨o、L. Prediger、A. Honkela 和 S. Kaski。DPVIM：改进的差分隐私变分推断。arXiv 预印本 arXiv:2210.15961，2022年。A. J ohnson、L. Bulgarelli、T. Pollard、S. Horng、L. A. Celi 和 R. Mark。MIMIC-IV。PhysioNet。在线获取：https://physionet.org/content/mimiciv/1.0/（访问日期：2021年8月23日），2020年。J. Jordon、J. Yoon 和 M. Van Der Schaar。PATE-GAN：通过差分隐私保证生成合成数据。载于《国际学习表征会议》，2018年。V. Karwa、D. Kifer 和 A. B. Slavkovi´c。基于变分近似的私有后验分布。arXiv 预印本 arXiv:1511.07896，2015年。D. P. Kingma 和 J. Ba。Adam：一种随机优化方法。arXiv 预印本 arXiv:1412.6980，2014年。

D. P. Kingma 和 P. Dhariwal。GLOW：采用可逆 1×1 卷积的生成流。神经信息处理系统进展，31，2018年。D. P. Kingma 和 M. Welling。自编码变分贝叶斯。*arXiv* 预印本 *arXiv:1312.6114*，2013年。D. P. Kingma、T. Salimans、R. Jozefowicz、X. Chen、I. Sutskever 和 M. Welling。利用逆自回归流改进变分推断。神经信息处理系统进展，29:4743–4751，2016年。M. Kroll。局部差分隐私下固定点的密度估计。电子统计学杂志，15(1):1783–1813，2021。N. Li、T. Li 和 S. Venkatasubramanian。t-接近性：超越 k-匿名性和 l-多样性的隐私保护。载于*2007*年 *IEEE* 第*23*届国际数据工程会议，第 106–115 页。IEEE，2007年。F. Liu。基于模型的差分隐私数据合成，以及对多重合成差分隐私数据的统计推断。数据隐私汇刊，15(3):141–175，2022年。A. Maceira、S. Prasad、M. Khan 和 D. Pennell。通过稳态自由进动心血管磁共振测得的正常左心室收缩和舒张功能。心血管磁共振杂志，8(3):417–426，2006年。I. Mironov。Rényi 差分隐私。载于*2017*年 *IEEE* 第*30*届计算机安全基础研讨会（*CSF*），第 263–275 页。IEEE，2017年。A. Narayanan 和 V. Shmatikov。大型稀疏数据集的稳健去匿名化。*IEEE* 安全与隐私研讨会，第 111–125 页，2008年。G. Papamakarios、T. Pavlakou 和 I. Murray。用于密度估计的掩码自回归流。神经信息处理系统进展，30，2017年。N. Papernot、M. Abadi、U. Erlingsson、I. Goodfellow 和 K. Talwar。利用私有训练数据进行深度学习的半监督知识迁移。*arXiv* 预印本 *arXiv:1610.05755*，2016 年。´N. Papernot 、 S. Song 、 I. Mironov 、 A. Raghunathan 、 K. Talwar 和 U. Erlingsson。使用 PATE 实现可扩展的隐私保护学习。*arXiv* 预印本 *arXiv:1802.08908*，2018年。M. Park、J. Foulds、K. Chaudhuri 和 M. Welling。隐私设置下的变分贝叶斯（VIPS）。*arXiv* 预印本 *arXiv:1611.00340*，2016年。B. Pfitzner 和 B. Arnrich。DPD-FVAE：使用带差分隐私解码器的联邦变分自编码器生成合成数据。*arXiv* 预印本 *arXiv:2211.11591*，2022年。T. J. Pollard、A. E. Johnson、J. D. Raffa、L. A. Celi、R. G. Mark 和 O. Badawi。eICU 协作研究数据库：一个可免费使用的多中心重症护理研究数据库。《科学数据》，5(1):1–13，2018年。D. Rankin、M. Black、R. Bond、J. Wallace、M. Mulvenna、G. Epelde 等。医疗保健中使用合成数据进行监督式机器学习的可靠性：用于保护数据共享隐私的模型。*JMIR* 医学信息学，8(7):e18910，2020年。D. Rezende 和 S. Mohamed。使用归一化流进行变分推断。载于国际机器学习会议，第 1530–1538 页。PMLR，2015年。

D. B. Rubin. 统计披露限制。《官方统计学杂志》，9(2):461–468，1993年。P. Samarati 和 L. Sweeney. 披露信息时如何保护隐私：k-匿名性及其通过泛化和抑制来实施的方法。技术报告，*SRI International*，1998年。M. Sharma、M. Hutchinson、S. Swaroop、A. Honkela 和 R. E. Turner. 差分隐私联邦变分推断。*arXiv* 预印本 *arXiv:1911.10563*，2019年。R. Shokri、M. Stronati、C. Song 和 V. Shmatikov. 针对机器学习模型的成员推断攻击。载于 *2017*年 *IEEE* 安全与隐私研讨会（*SP*），第3–18页。IEEE，2017年。G. Simonneau、M. Gatzoulis、I. Adatia、C. Celermajer、D. 和 Denton、A. Ghofrani、M. Sanchez Gomez、R. Kumar、M. Landzberg、R. Machado、H. Olschewski、I. Robbins 和 S. R. 肺动脉高压临床分类更新。《美国心脏病学会杂志》，62(25 Supplement):D34–D41，2013年。L. Sweeney. 共享医疗数据时维护患者保密性，需要技术与政策之间的协同关系。麻省理工学院人工智能实验室，AIWP-WP344，1997年。L. Sweeney. 在华盛顿州数据中将已知患者与健康记录匹配。社会科学研究网络，2013年。id=2289850。SHARED 团队。合成健康与研究数据（SHARED）项目，2023年。网址：https://shared.landen.co/。访问日期：2023年1月8日。T. Tieleman、G. Hinton 等。讲义 6.5-RMSProp：将梯度除以其近期幅度的移动平均值。*COURSERA*：机器学习神经网络，4(2):26–31，2012年。S. Van Buuren 和 K. Groothuis-Oudshoorn. mice：R 中基于链式方程的多元插补。《统计软件杂志》，45:1–67，2011年。C. Waites 和 R. Cummings. 用于隐私保护密度估计的差分隐私归一化流。*arXiv* 预印本 *arXiv:2103.14068*，2021年。L. Wasserman 和 S. Zhou. 差分隐私的统计框架。《美国统计学会杂志》，105(489):375–389，2010年。L. Xie、K. Lin、S. Wang、F. Wang 和 J. Zhou. 差分隐私生成对抗网络。*arXiv* 预印本 *arXiv:1802.06739*，2018年。A. Yale、S. Dash、R. Dutta、I. Guyon、A. Pavao 和 K. P. Bennett. 隐私保护合成健康数据的生成与评估。《神经计算》，416:244–255，2020年。K. Yared、P. Noseworthy、A. Weyman、E. McCabe、M. Picard 和 A. Baggish. 经胸超声心动图检查中，肺动脉加速时间可准确估算肺动脉收缩压。《美国超声心动图学会杂志》，24

(6)：687–692 ，2011年。