差分隐私算法基础

摘要

隐私保护数据分析的问题具有跨越多个学科的悠久历史。 随着关于个人的电子数据变得越来越详细，并且随着技术实现对这些数据的更强大的收集和管理，对于强大，有意义和数学上严格的隐私的定义的需求增加，以及满足此定义的计算丰富的算法类别。差分隐私就是这样一个定义。

在激励和讨论差分隐私的含义之后，这本专著的优势在于使用查询 - 释放问题作为一个持续的例子，致力于实现差分隐私的基本技术，以及这些技术在创意组合中的应用。关键点在于，通过重新思考计算目标，人们通常可以获得比通过差分隐私实现有条不紊地替换非私有计算的每个步骤所实现的结果更好的结果。尽管有一些非常强大的计算结果，但仍存在一些根本的局限性 - 不仅仅是关于可以通过差分隐私能够实现什么，而且通过任何可以防止完全隐私破坏的方法能够实现什么。实际上，这里讨论的所有算法都保持预防任意计算能力的对手的差分隐私。某些算法是计算密集型的，其他算法则是高效的。讨论了对手和算法的计算复杂性。

然后，我们从基础知识转向查询发布以外的应用程序，讨论机制设计和机器学习的差分隐私方法。 关于差分隐私算法的绝大多数文献都考虑了一个单一的静态数据库，该数据库需要经过许多分析。 讨论了其他模型中的差分隐私，包括分布式数据库和数据流计算。

最后，我们注意到这项工作是对差分隐私问题和技术的全面介绍，但并非旨在进行详尽的调查 - 目前在差分隐私方面有大量工作，我们只能涵盖 它的一小部分。

前言

隐私保护数据分析的问题具有跨越多个学科的悠久历史。 随着关于个人的电子数据变得越来越详细，并且随着技术实现对这些数据的更强大的收集和管理，对于强大，有意义和数学上严格的隐私的定义的需求增加，以及满足此定义的计算丰富的算法类别。差分隐私就是这样一个定义。

在激励和讨论差分隐私的含义之后，这本专著的优势在于使用查询 - 释放问题作为一个持续的例子，致力于实现差分隐私的基本技术，以及这些技术在创意组合的应用（第3-7节）。关键点在于，通过重新思考计算目标，人们通常可以获得比通过差分隐私实现有条不紊地替换非私有计算的每个步骤所实现的结果更好的结果。

尽管有一些非常强大的计算结果，但仍存在一些根本的局限性 - 不仅仅是关于可以通过差分隐私能够实现什么，而且通过任何可以防止完全隐私破坏的方法能够实现什么（第8节）。

实际上，第10和11节这里讨论的所有算法都保持预防任意计算能力的对手的差分隐私。某些算法是计算密集型的，其他算法则是高效的。第9节讨论了对手和算法的计算复杂性。

然后，我们从基础知识转向查询发布以外的应用程序，讨论机制设计和机器学习的差分隐私方法。关于差分隐私算法的绝大多数文献都考虑了一个单一的静态数据库，该数据库需要经过许多分析。第12节讨论了其他模型中的差分隐私，包括分布式数据库和数据流计算。

最后，我们注意到这项工作是对差分隐私问题和技术的全面介绍，但并非旨在进行详尽的调查 - 目前在差分隐私方面有大量工作，我们只能涵盖 它的一小部分。

1.差分隐私的承诺

“差分隐私”描述了数据持有者或*curator*对数据主体的承诺：“允许您的数据用于任何研究或分析中，在无论其他什么研究、数据集或信息源都可以获得的情况下，您不会受到对手或其他任何人的影响。”在最好的情况下，差分隐私数据库机制可以使机密数据广泛用于准确的数据分析，而无需借助数据清洁室，数据使用协议，数据保护计划或受限制的视图。 尽管如此，最终将消耗数据效用：信息恢复基本法规定，对许多问题的过于准确的回答将以惊人的方式破坏隐私。[[1]](#footnote-1)差分隐私算法研究的目标是尽可能地推迟这种必然性。

差分隐私解决了在学习有关人口的有用信息的同时不学习个人的悖论。 医疗数据库可能会告诉我们，吸烟会导致癌症，这会影响保险公司对吸烟者长期医疗费用的看法。 吸烟者是否受到分析的伤害？ 也许 - 如果保险公司知道他吸烟，他的保险费可能会上涨。 他也可能得到帮助 - 了解他的健康风险，他进入戒烟计划。 吸烟者的隐私受到了损害吗？ 研究后对他的了解肯定比之前多，但他的信息是否“泄露”了？差分隐私将认为它不是，理由是对吸烟者的影响是相同的，与他是否在研究中无关。研究中得出的结论影响吸烟者，而不是他在数据集中的存在与否。

差分隐私确保将达到相同的结论，例如吸烟导致癌症这条结论与是否有人选择进入或退出数据集无关。具体而言，它确保任何输出序列（对查询的响应）“基本上”同样可能发生，而与任何个体的存在与否无关。这里，概率取自隐私机制（由数据curator控制的事物）做出的随机选择，术语“基本上”由参数捕获。较小的ϵ将产生更好的隐私（并且不太准确的响应）。

差分隐私是一种*定义*，而不是算法。对于给定的计算任务T和给定的ϵ值，将存在许多差分隐私算法，用于以ϵ-差分隐私方式实现T.有些人比其他人有更好的准确性。当ϵ很小时，为T找到高度准确的ϵ-差分隐私算法可能很困难，就像为特定计算任务找到数值稳定的算法可能需要付出努力一样。

1.1隐私保护数据分析

差分隐私是针对隐私保护数据分析问题量身定制的隐私定义。我们简要介绍了解决此问题的其他方法的一些问题。

*数据无法完全匿名并保持有用*。一般来说，数据越丰富，它就越有趣和有用。这导致了“匿名化”和“删除个人身份信息”的概念，其中希望可以抑制部分数据记录并将其余部分发布并用于分析。然而，数据的丰富性使得能够通过有时令人惊讶的字段或属性的集合来“命名”个体，例如邮政编码，出生日期和性别的组合，甚至三部电影的名称和观看了这些电影的大致日期。此“命名”功能可用于*链接攻击（linkage attack）*，以将“匿名”记录与不同数据集中的非匿名记录进行匹配。因此，马萨诸塞州州长的医疗记录通过对比匿名医疗事故数据与（公众可用的）选民登记记录被识别出来，以及Netflix订阅用户通过与互联网电影数据库（IMDb）链接被识别出来，这些用户的观看历史包含在被Netflix发布出来作为推荐算法比赛的训练数据的匿名电影记录集合中。

差分隐私使链接攻击失效（linkage attack）：由于差分隐私是数据访问机制的属性，并且与攻击者可用的辅助信息的存在与否无关，访问IMDb将不再允许历史记录的在Netflix训练集中的人比不在训练集中的人更容易受到链接攻击（linkage attack）。

*重新识别“匿名”记录不是唯一的风险*。重新识别“匿名”数据记录显然是不可取的，这不仅是因为重新识别本身，这肯定揭示了数据集中的成员身份，而且因为记录可能包含易暴露的信息，如果与个人联系起来，可能会造成伤害。在给定日期来自特定紧急护理中心的医疗事故记录的集合可以仅列出少量不同的投诉或诊断。邻居在相关日期访问机构的附加信息为邻居的状况的可能诊断提供了相当窄范围。可能无法将特定记录与邻居匹配的事实为邻居提供最小的隐私保护。

*对大型集合的查询不具有保护性*。关于特定个体的问题不能被准确地安全地回答，并且实际上人们可能希望拒绝它们（如果在计算上可以识别它们）。强制查询超过大集并不是灵丹妙药，如下面的差异攻击所示。假设已知X先生在某个医疗数据库中。总的来说，两个大问题的答案“数据库中有多少人具有镰状细胞特征？”和“数据库中有多少人，而不是X，具有镰状细胞特征？”产生镰状细胞状态X先生

*查询审核是有问题的*。人们可能会试图审核查询和响应的顺序，目的是在根据历史记录，回答当前查询会破坏隐私的情况下阻止任何响应。例如，审核员可能正在寻找构成差异攻击的查询对。这种方法有两个困难。首先，*拒绝*回答查询本身是有可能披露的。其次，查询审计在计算上是不可行的;实际上，如果查询语言足够丰富，甚至可能不存在用于判定一对查询是否构成差分攻击的算法过程。

*摘要统计不是“安全的”*。从某种意义上说，摘要统计作为隐私解决方案概念的失败可以立即从刚才描述的差异攻击中获得。摘要统计的其他问题包括针对数据库的各种重建攻击，其中每个人具有要保护的“秘密位”。实用目标可能是允许例如“满足属性P的人有多少秘密比特值1？”这一形式的问题。另一方面，对手的目标是显着增加他猜测个人秘密比特位的机会。第8.1节中描述的重建攻击表明即使是这种类型的线性数量的查询也难以保护：除非引入足够的不准确性，否则几乎所有的秘密比特都可以被重建。

发布汇总统计数据的风险的一个引人注目的例子是应用统计技术，最初用于确认或驳斥个体DNA在法医学混合物中的存在，以统治个体进出全基因组关联研究。根据人类基因组计划的网站，“单核苷酸多态性或SNP（发音为”snips”）是发生在基因组序列中单个核苷酸（A，T，C或G）改变时的DNA序列变异。例如，SNP可能会将DNA序列AAGGCTAA改为ATGGCTAA。“在这种情况下，我们说有两个等位基因：A和T.对于这样的SNP，我们可以询问，给定一个特定的参考群体，两个可能的等位基因的频率分别是多少？考虑到参考群体中SNP的等位基因频率，我们可以检查这些频率对于具有特定疾病的亚群（“病例”组）如何不同，寻找与疾病相关的等位基因。出于这个原因，全基因组关联研究可能包含大量SNP的病例组的等位基因频率。根据定义，这些等位基因频率只是汇总统计数据，而（错误的）假设是，凭借这种聚合，它们可以保护隐私。然而，考虑到个体的基因组数据，理论上可以确定个体是否属于病例组（并且因此具有疾病）。作为回应，美国国立卫生研究院和惠康国际信托基金会终止公众查阅他们资助的研究中的汇总频率数据。

即使对于差别隐私，这也是一个具有挑战性的问题，因为涉及的测量数量很大 - 几十万甚至一百万 - 而且任何情况下的个人数量相对较少。

*“普通”事实不是“好的”*。如果随着时间的推移跟踪数据主体，揭示“普通”事实，例如购买面包，可能会有问题。例如，考虑T先生，他经常年复一年地买面包，直到突然转向很少买面包。分析师可能会得出结论，T先生最有可能被诊断出患有2型糖尿病。分析师可能是正确的，或者可能是不正确的;无论哪种方式，T先生都受到伤害。

“*只是少数*。”在某些情况下，特定技术实际上可以为数据集的“典型”成员或更一般地“最典型”成员提供隐私保护。在这种情况下，人们常常听到技术充分的论点，因为它会损害“只有少数”参与者的隐私。撇开极少的那部分人可能恰恰对他们来说隐私是最重要的担忧，“只是少数”的哲学本质上并非没有价值：要做出社会判断，权衡成本和利益。尚未制定出与“少数”哲学相一致的明确的隐私定义;然而，对于单个数据集，可以通过随机选择行的子集并完整地释放它们来实现“仅仅几个”隐私（引理4.3，第4节）。描述可以在随机子样本上执行的统计分析质量的采样边界决定了要释放的行数。当“仅少数”理念被拒绝时，差分隐私提供了另一种选择。

1.2书目笔记

Sweeney [81]将选民登记记录与“匿名”医疗遭遇数据联系起来; Narayanan和Shmatikov对Netflix发布的匿名排名数据进行了linkage attack [65]。有关法医学混合存在的工作归功于荷马等人[46]。第一次重建袭击是由于Dinur和Nissim [18]。

2 基本条款

本节开始介绍差分隐私的正式定义，并列举其一些关键属性。

2.1计算模型

我们假设存在一个值得信赖并且值得信赖的*curator*，他将*个人*数据保存在数据库D中，通常由n行组成。直觉是每行包含单个人的数据，并且，直观地说，隐私目标是同时在允许对整个数据库进行统计分析的时候，同时保护每一行个人数据。

在*非交互式*或*离线*模型中，curator会产生一些对象，例如“synthetic database”，汇总统计信息的集合，或者“sanitized database”，一劳永逸。在此版本发布后，curator不再发挥作用，原始数据可能会被销毁。

*查询*是应用于数据库的函数。*交互式*或*在线*模型允许数据分析员自适应地询问查询，基于观察到的对先前查询的响应来决定接下来要提出的查询。

可信赖的curator可以由一组个人运行的协议替换，(这协议?)使用用于安全多方协议的加密技术，但在大多数情况下，我们对加密假设没兴趣。第12节描述了文献中研究的这个和其他模型。

当事先知道所有查询时，非交互模型应该给出最佳准确度，因为它能够在知道查询结构的情况下关联噪声。相反，当事先不知道关于查询的信息时，非交互模型提出了严峻的挑战，因为它必须为所有可能查询提供答案。正如我们将要看到的，为了确保隐私，甚至防止隐私灾难，准确性必然会随着提出的问题数量而恶化，并且为所有可能问题提供准确答案将是不可行的。

*隐私机制*或简称*机制*是一种算法，它将数据库，a universe X of data types（所有可能的数据库行的集合），随机位以及可选的一组查询作为输入，并生成输出字符串。希望输出字符串可以被解码以产生相对准确的查询答案（如果存在后者）。如果没有呈现查询，那么我们处于非交互式情况，并且希望可以解释输出字符串以提供对未来查询的答案。

在某些情况下，我们可能要求输出字符串是*合成数据库*。这是从可能的数据库行的Universe X中提取的多集。在这种情况下，解码方法是对合成数据库执行查询，然后应用某种简单的变换，例如乘以缩放因子，以获得对查询的真实答案的近似。

2.2定义私人数据分析

在数据分析的背景下定义隐私的一种自然方法是要求分析人员在分析完成后对数据集中的任何个人的了解，不超过分析开始前的了解。通过要求对手关于个人的前后观点（即，在访问数据库之前和之后）不应该“太不同”，或者访问数据库不应该使对手对任何个人的看法改变“太多”，将这个目标形式化也是很自然的。然而，如果数据库教授任何东西，这种隐私概念是无法实现的。例如，假设对手的（不正确的）先前观点是每个人都有2个左脚。访问统计数据库得到回答几乎每个人都有一个左脚和一个右脚。对手现在对于任何给定的响应者是否有两个左脚有一个非常不同的看法。

对于定义隐私的前/后或“没有学到”方法的部分吸引力是如果没有了解到任何关于个体的知识，那么个人就不会受到分析的伤害。然而，“吸烟导致癌症”的例子表明这种直觉是有缺陷的;罪魁祸首是辅助信息（X先生抽烟）。

定义隐私的“没有学到的”方法让人联想到密码系统的语义安全性。粗略地说，语义安全性表明没有从密文中了解到明文（未加密的消息）的信息。也就是说，在看到密文之后知道明文的任何事情都是看密文前已知的。因此，如果有辅助信息，比如密文是 “狗”或“猫”的加密，则密文不会泄漏关于“狗”或“猫”中的哪一个已被加密的进一步信息。形式上，这是通过比较窃听者猜测“狗”和“猫”那个被加密的能力与所谓的对手模拟器猜测同样的事情的能力进行建模的，对手模拟器具有辅助信息但无法访问密文。如果对于每个窃听对手，以及所有辅助信息（对手和模拟器都是秘密的），对手模拟器与窃听者具有基本相同的猜测几率，则系统享有语义安全性。当然，为了使系统有用，合法的接收器必须能够正确地解密消息;否则语义安全可以通过简单的方式实现。

我们知道，在标准计算假设下，存在语义安全的密码系统，那么为什么我们不能建立语义安全的私有数据库机制来产生查询答案，同时保持个别行的秘密？

首先，类比并不完美：在语义上安全的密码系统中有三个方：消息发送者（加密明文消息），消息接收者（解密密文）和窃听者（因为无法了解在发送之前她还不知道的明文而感到沮丧）。相比之下，在私人数据分析的设置中，只有两个方面：curator，它运行隐私机制（类似于发送者）和数据分析员，他们接收对查询的信息响应（如消息接收者），也试图挤出有关个人的泄露隐私的信息（如窃听者）。因为合法接收者与窥探对手是同一方，所以加密的类比是有缺陷的：拒绝向对手发送所有信息意味着拒绝向数据分析员提供所有信息。

其次，与加密方案一样，我们要求隐私机制有用，这意味着它教会分析师以前不知道的东西。这种教学对于对手模拟器是不可用的;也就是说，没有模拟器可以“预测”分析师学到了什么。因此，我们可以将数据库视为随机（不可预测）位的弱源，我们可以从中提取一些非常高质量的随机性以用作*random pad*。这可以用在加密技术中，其中秘密消息被添加到随机值（“random pad”）以便产生信息理论上隐藏秘密的字符串。只有知道random pad的人才能知道这个秘密;无论他或她的计算能力如何，对pad一无所知的任何一方都不了解这个秘密。在给定数据库访问权限的情况下，分析师可以学习random pad，但是没有访问数据库的对手模拟器根本无法学习pad。因此，作为辅助信息，使用random pad秘密的密文，分析师可以解密秘密，但是对手模拟器根本不知道秘密是什么。这导致**对手/分析师**学习秘密的能力与**对手模拟器**做同样事情的能力之间存在巨大差异，消除了任何与语义安全无关的任何希望。

吸烟导致癌症的例子和对语义安全的希望，；两者的障碍都是辅助信息。显然，为了有意义，即使在“合理的”辅助知识的背景下，隐私保证也必须保持，但是将合理的辅助知识与任意辅助知识分开是有问题的。例如，使用政府数据库的分析师可能是大型搜索引擎公司的雇员。关于这样一个人可获得的辅助知识信息的“合理”假设是什么？

2.3形式化差分隐私

我们将从差分隐私的技术定义开始，然后继续解释它。差异化隐私将通过*流程*提供隐私;特别是它会引入随机性。通过随机过程进行隐私的早期例子是*随机响应*，这是一种在社会科学中研究的技术，用于收集关于尴尬或非法行为的统计信息，随机响应通过属性P来获取。要求研究参与者报告他们是否有属性P，如下所示：

1.抛一枚硬币。

2.如果是反面，那就回应如实。

3.如果是正面，则抛第二枚硬币，如果是正面则回答“是”，如果是反面则回答“否”。

“隐私”来自任何结果的合理否认;特别是，如果属性P对应于从事非法行为，即使是“是”答案也不是有罪的，因为无论受访者是否确实拥有属性P，出现这个答案的概率至少为1/4。准确性来自于理解噪声生成程序（从随机化中引入虚假的“是”和“否”答案）：“是”答案的预期数量是没有P属性的参与者数量的1/4，加上有P属性的参与者数量的3/4。因此，如果p是具有属性P的参与者的真实分数，则预期的“是”答案的数量是（1/4）（1-p）+（3/4）p =（1 / 4）+p/ 2。因此，我们可以将p估计为回答“是”的分数的两倍减去1/2，即2（（1/4）+ p / 2） - 1/2。

随机化是必不可少的;更准确地说，任何非平凡的隐私保证，无论是现在的甚至未来的辅助信息来源，包括其他数据库，研究，网站，在线社区，八卦，报纸，政府统计等，都需要随机化。这是一个简单的混合论证，我们现在草拟。假设为了矛盾，我们有一个非平凡的确定性算法。非平凡性表示存在一个查询和两个在此查询下产生不同输出的数据库。我们一次更改一行，我们看到存在一对数据库，这些数据库的区别仅在于一行的值，同一查询会产生不同的输出。知道数据库是这两个几乎相同的数据库之一的对手学习未知行中数据的值。

因此，我们需要讨论随机算法的输入和输出空间。在整篇专着中，我们使用离散概率空间。有时我们会将算法描述为从连续分布中采样，但这些算法应始终以适当谨慎的方式离散化为有限精度（参见下面的备注2.1）。通常，具有域A和（离散）范围B的随机算法将与从A到B上的概率单纯形(Probability Simplex)的映射相关联，表示为：

**定义2.1（概率单纯形 Probability Simplex）**。 给定离散集B，B上的*Probability Simplex*，表示为，定义为：

**定义2.2（随机算法）**。 具有域A和离散范围B的随机算法M与映射相关联。 在输入a∈A时，算法M对于每个b∈B以概率输出。概率空间在算法M的硬币翻转上。

我们将数据库x视为来自宇宙X的记录集合。通过直方图表示数据库通常很方便： ，其中每个条目表示数据库x中类型为i∈X的元素数量(记录会重复出现么？)（我们略微滥用符号，让符号N表示所有非负整数的集合，包括零）。 在这个表示中，两个数据库x和y之间距离的自然度量将是它们的距离：

**定义2.3（数据库之间的距离）**。数据库x的范数表示为并定义为：

两个数据库x和y之间的距离是

注意是数据库x的大小（即，它包含的记录数）的度量，并且

是x和y之间有多少记录不同的度量。

数据库也可以由行（X的元素）的多个集或甚至行的有序列表表示，这是集合的特殊情况，其中行号成为元素名称的一部分。在这种情况下，数据库之间的距离通常由汉明距离测量，即它们不同的行数。

但是，除非另有说明，否则我们将使用上述直方图表示。（但是，请注意，即使直方图符号在数学上更方便，在实际实现中，多集表示通常也会更加简洁）。

我们现在准备正式定义差分隐私，这直观地保证了随机算法在相似似输入数据库上的行为类似。

**定义2.4（差分隐私）**。 具有域的随机算法M是的条件是对于所有S⊆Range(M)和对于所有x，y∈ 其中

,

其中概率空间是M机制抛硬币的概率空间。如果，我们说M是差分隐私的。

通常我们感兴趣的是值小于数据库大小中任何多项式的倒数。特别是，的值约为是非常危险的：他们允许通过发布少数数据库参与者的完整记录来“保护隐私” - 正是第1节中讨论的“仅仅是少数”哲学。

然而，即使当可忽略不计时，和差分隐私之间存在理论上的区别。其中最重要的是量化顺序的转换。差分隐私确保对于机制的每次运行，观察到的输出（几乎）同样可能同时在每个相邻数据库上观到。相比之下-差分隐私说对于每对相邻数据库x、y，事后观察值由数据库x或数据库y产生的可能性相差太大的情况是不可能出现的。然而，给定输出，有可能找到数据库y，使得在数据库y上产生比在数据库x上产生的可能性更大。也就是说，分布中的的质量可以显着大于其在分布中的质量。

数量

对我们很重要;我们将其称为观察所引起的*隐私损失*。这种损失可能是正面的（当事件更可能在x下而不是在y下发生）或者可能是负面的（当事件更可能在y下而不是在x下发生）。正如我们将在引理3.17中看到的那样，差分隐私确保对于所有相邻的x、y，隐私损失的绝对值将以为界，概率至少为。与往常一样，概率空间是M机制抛硬币的概率空间。

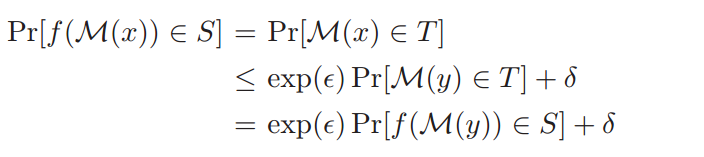
差分隐私不受*后处理*的影响：数据分析师在没有关于私有数据库的额外知识的情况下，无法计算私有算法M的输出函数并使其成为不太差分隐私。也就是说，如果算法保护个人的隐私，那么数据分析师就无法在正式定义下甚至在任何直观意义上增加隐私损失 - 只需坐在角落里并考虑算法的输出。形式上，具有差分隐私算法M的数据无关映射f的组成也是差分隐私的：

**命题2.1（后处理）**。设M：→R是差分隐私的随机算法。设是一个随意的随机映射。然后是差分隐私的(f(g(x))还是g(f(x)))。

*证明*。我们证明了确定性函数 条件下的命题。然后得出证明结果，因为任何随机映射都可以分解为确定性函数的凸组合，并且差分隐私机制的凸组合是差分隐私的。

固定任何一对相邻数据库x、y并且，并固定任何事件

。设。然后我们有：



这是我们想要的。

差分隐私以直接的方式组成遵循定义2.4：两个差分隐私机制的组成是差分隐私。更一般地说（定理3.16），“epsilons和deltatas加起来”：k个差分隐私机制的组成是差分隐私的,其中第i个机制是差分隐私(1≤i≤k)。

差分隐私机制的组隐私也立即遵循定义2.4，隐私保证的强度随着组的大小线性下降。

**定理2.2。**对于大小为k的组，任何差分隐私机制M是差分隐私。也就是说，对于所有且所有S⊆Range(M)



其中概率空间是M机制抛硬币的概率空间。

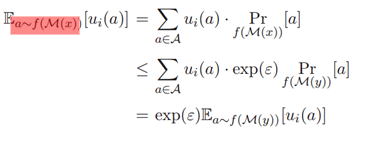
例如，这解决了包括多个家庭成员的调查中的隐私问题。[[2]](#footnote-2)

更一般地说，构成和群体隐私不是一回事，第3.5.2节（定理3.20）中的改进的组合界限改进了k因子，做不到 - 并且不能 - 为群体隐私产生相同的收益，即使。

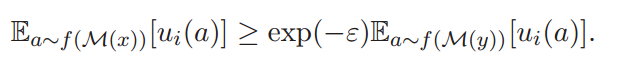
2.3.1差分隐私的承诺

**经济观。**差分隐私承诺保护个人不受由于其数据在私有数据库X中而可能面临的任何额外伤害，如果他们的数据不是X的一部分，他们就不会面临这些伤害。虽然一旦差分隐私机制M的结果M（x）被发布，个人可能确实面临伤害，但差别隐私承诺，他们被伤害的可能性并未因选择参与而显着增加。这是一个非常实用的隐私定义，因为当一个人决定是否将她的数据包含在将以差分隐私方式使用的数据库中时，她正在考虑的正是这种差异：假设她参与受到伤害的概率与假设她不参与受到伤害的概率比较。她无法控制数据库的其余容。鉴于差分隐私的承诺，她确信，从未来的伤害角度来看，她应该参与与不参与之间几乎无关紧要。鉴于任何激励 - 从利他主义到金钱奖励 - 差分隐私可能说服她允许使用她的数据。这种直觉可以在效用理论意义上形式化，我们在这里简要地描述。

考虑一个对所有可能的未来事件的集合有任意偏好的个体i，我们用A表示未来事件。这些偏好由效用函数表示，我们说个体在即将到来的事件上的体验效用。假设是包含个体is私有数据的数据集，并且M是差分隐私算法。设y是与x相同的数据集，除了它不包括个体i的数据（具体而言，），并且令是决定未来事件A分布的（任意）函数，以机制M的输出为条件。通过保证差别隐私，以及对命题2.1保证的任意后处理的弹性，我们得到：



同理，



因此，通过承诺保证差别隐私，数据分析师可以向个人承诺，他的预期未来效用不会受到超过因子的伤害。请注意，此承诺*独立*于个人*is*实用函数，并且同时适用于可能具有完全不同的实用函数的多个人。

2.3.2差分隐私不承诺什么

正如我们在吸烟导致癌症中看到的那样，虽然差别隐私是一种极其有力的保障，但它并不保证无条件免受伤害。它也不会创造以前根本不存在的隐私。更一般地说，差别隐私并不能保证一个人认为是秘密的东西将保密。它只是确保参与调查本身不会被披露，参与也不会导致披露任何有助于调查的细节。从调查中得出的结论很可能反映出个人的统计信息。旨在发现特定疾病的早期指标的健康调查可能会产生强有力的，甚至是决定性的结果;这些结论适用于特定个人并不是破坏隐私权差异的证据;个人甚至可能没有参与调查（同样，差别隐私确保无论个人是否参与调查，都可以非常相似的概率获得这些结论性结果）。特别是，如果调查告诉我们特定的私人属性与公开的可观察属性强烈相关，这并不违反差别隐私，因为无论是否有任何受访者，都会以几乎相同的概率观察到相同的相关性。

**差别隐私的定性属性**。介绍并正式定义差分隐私后，我们重新考虑其关键的理想品质。

1.*防范任意风险*，超越重新识别保护范围。

2.*自动中和链接攻击*，包括所有过去，现在和未来数据集以及其他形式和

辅助信息来源的尝试。

3.*量化隐私损失*。差分隐私不是二元概念，并且具有隐私损失的度量。这

允许在不同技术之间进行比较：对于隐私损失的固定限制，哪种技术提

供更好的准确性？为了固定的准确性，哪种技术提供更好的隐私？

4.*组成*。也许最重要的是，损失的量化还允许分析和控制多次计算的累积

隐私损失。了解组合下差分隐私机制的行为，可以从简单的差分隐私构

建块中设计和分析复杂的差分隐私算法。

5.*团体隐私*。差异化隐私允许分析和控制群组（例如家庭）引起的隐私损

失。

6.*后处理下的闭包*。差分隐私不受后处理的影响：数据分析师在没有关于

私有数据库的额外知识的情况下，无法计算差分隐私算法M的输出函

数，并使其差异性较小。也就是说，无论是在正式定义下，还是在任何

直观意义上，数据分析师都无法通过坐在角落里思考算法的输出，无论

可用的辅助信息是什么，都不会增加隐私损失。

这些是差分隐私的信号属性。我们可以证明相反吗？也就是说，这些属性或其某些子集是否意味着差分隐私？差别隐私可以在这些方面被削弱并且仍然有意义吗？这些是开放性问题。

2.3.3关于定义的最后评论

**隐私的粒度**。应仔细审查差分隐私的声明，以确定承诺隐私的粒度级别。差分隐私承诺即使修改了数据库中的单个条目，算法的行为也将大致保持不变。但是什么构成了数据库中的单个条目？例如，考虑采用图形形式的数据库。这样的数据库可以编码社交网络：每个个体i∈[n]由图中的顶点表示，并且个体之间的友谊由边表示。

我们可以在与个体对应的粒度级别上考虑差分隐私：也就是说，我们可以要求差分隐私算法对从图形中添加或移除任何顶点不敏感。这提供了强大的隐私保障，但实际上可能比我们需要的更强。添加或删除单个顶点后，最多可以添加或删除图中的n个边。根据我们希望从图中学习的内容，对n个边缘移除的不敏感可能是一个不可能满足的约束条件。

另一方面，我们可以考虑与边缘对应的粒度级别的差分隐私，并要求我们的算法仅对图中单个或少量边缘的添加或删除不敏感。这当然是一个较弱的保证，但可能仍然足够用于某些目的。非正式地说，如果我们在单个边的水平上承诺差分隐私，那么没有数据分析师能够得出关于图中边的任何子集的存在的任何结论。在某些情况下，大量社交联系可能不被视为敏感信息：例如，个人可能觉得不需要隐瞒他的大多数联系人与他所在城市或工作场所的个人有关的事实，因为他住在哪里和他工作的地方是公共信息。另一方面，可能存在少量的社交联系，其存在是高度敏感的（例如，未来的新雇主，或亲密的朋友）。在这种情况下，边隐私应该足以保护敏感信息，同时仍然允许比顶点隐私更全面地分析数据。 Edge隐私将保护此类个人的敏感信息，前提是他的朋友少于。

作为另一个例子，差分隐私电影推荐系统可以被设计成在单个电影的“事件”级别保护训练集中的数据，隐藏任何单个电影的观看/评级但不隐藏个人对牛仔西部片或血腥的热情，或在个人的整个观看和评级历史的“用户”级别。

**所有小型Epsilons都是一样的**。当很小时，差分隐私断言，对于所有相邻数据库对x、y和所有输出o，对手无法根据观察o来区分哪个是真正的数据库。当小时，不能差分隐私不一定是惊人的 - 例如，机制可以是差分隐私的。具有不同但小的epsilons的隐私保证的性质是非常相似的。但是大的值是什么？不是差分隐私只是说存在相邻的数据库和输出o，根据数据库x或y得出观察o的概率比值是很大的。o的输出可能是非常不可能的（这通过差分隐私来解决）;数据库x和y可能是非常人为的，并且不太可能出现在“现实世界”中;对手可能没有正确的辅助信息来识别出已发生的暴露输出;或者可能对数据库知之甚少以确定其对称差异的值。因此，就像弱密码系统可能泄露任何事情，从消息的最不重要的位到完整的解密密钥，未能成为或差分隐私可能包含有效无意义的隐私违反到完整数据库的泄露。一个大型的epsilon以其自己的时尚而变大。

**一些额外的形式/定义。**除了数据库x之外，我们的隐私机制M通常会将一些辅助参数作为输入。例如，w可以指定数据库x上的查询或查询的集合。机制可以（分别）用或中的一些或所有查询的差分隐私近似进行响应。对于所有，我们说机制满足差分隐私，如果每个w，满足差分隐私。

可以包括在w中的参数的另一个例子是安全参数，以控制应该有多小。也就是说，对于所有，应该是差分隐私的。通常，在整篇专论中，我们要求是中可忽略不计的函数，即。因此，我们认为在密码学上很小，而通常被认为是一个适度小的常数。

在辅助参数w指定查询的集合的情况下，我们将机制M称为*概要生成器*。概要生成器输出（差分隐私）概要A，其可用于计算中所有查询的答案。也就是说，我们要求存在重建过程R，使得对于每个输入v指定查询的，重构过程输出。通常，我们将要求具有高概率M产生概要A,使得使用A的重建过程计算准确的答案。也就是说，对于查询的所有或大多数（通过某种分布加权），误差将是有限的。我们偶尔会滥用符号并参考重构过程，将实际查询q（而不是它的某些表示v）作为输入，并输出。

概要的一个特例是*合成数据库*(synthetic database)。顾名思义，合成数据库的行与原始数据库的行具有相同的类型。合成数据库的一个优点是可以使用分析人员在原始数据库上使用的相同软件对它们进行分析，从而避免了对特殊重建程序R的需要。

**备注2.1。**由于浮点数实现的细微之处，在编写实数值机制（如拉普拉斯机制）时必须非常小心。否则，可以破坏差别隐私，因为在数据库x上具有非零概率的输出可能由于四舍五入而在相邻数据库y上具有零概率。这只是浮点的实现需要在差分隐私的背景下进行详细审查的一种方式，并且它不是唯一的。

2.4书目笔记

差分隐私的定义归功于Dwork等人[23];这里和文献中使用的精确式子首先出现在[20]中，归功于Dwork和McSherry。 “差分隐私”一词是由Michael Schroeder创造的。语义安全的不可能性归功于Dwork和Naor [25]。差分隐私机制的组成和群体隐私首先在[23]中得到解决。 差分隐私的组成首先在[21]中得到解决（但由于Dwork和Lei [22]，见附录B中的校正证明）。 Mironov提出了差分隐私对浮点数不适当实现的脆弱性，他提出了缓解措施[63]。

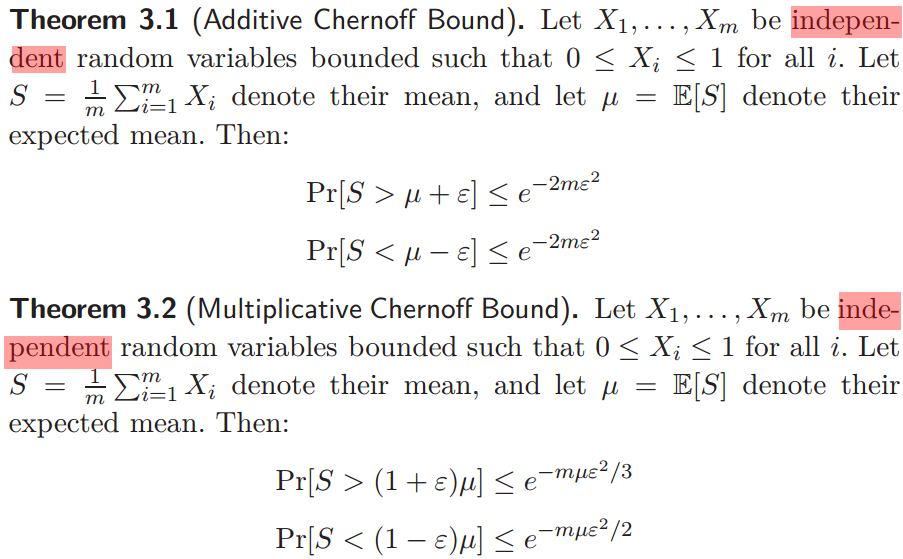
3 基本技术和组成定理

在回顾了一些概率工具之后，我们提出了拉普拉斯机制，它为实（矢量）值查询提供了差分隐私。这一点的应用自然引出了指数机制，这是一种从离散候选输出集中进行差分隐私选择的方法。 然后，我们分析了组成多个差分隐私机制所引起的累积隐私损失。最后，我们给出了一种方法--稀疏向量技术--用于私下报告可能非常大量计算的结果，前提是只有少数是“重要的”。

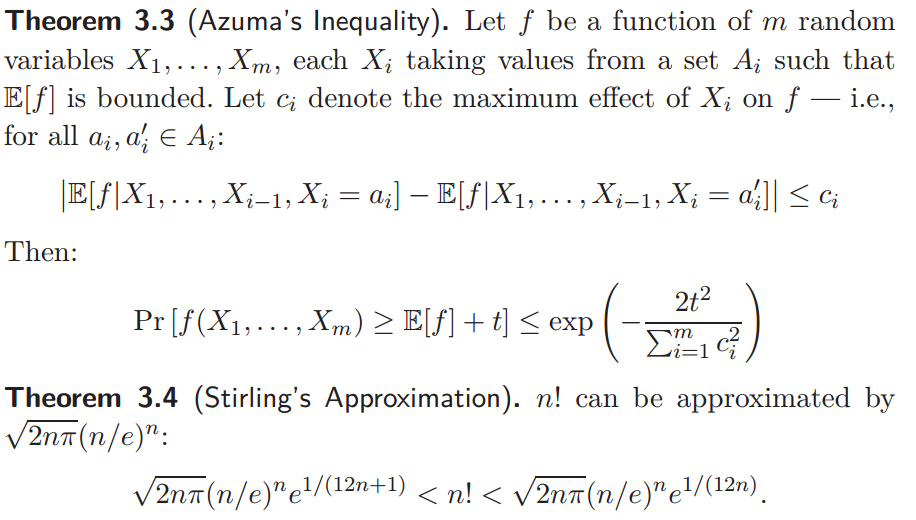
在本节中，我们将描述差分隐私中的一些最基本的技术，我们将一次又一次地使用。这里描述的技术构成了我们将要开发的所有其他算法的基本构块。

3.1有用的概率工具

以下集中不等式concentration inequalities通常是有用的。 我们以易于使用的形式陈述它们，而不是以最强的形式。



当我们**没有独立**的随机变量时，一切都不会丢失。我们仍然可以应用Azuma不等式：



3.2随机响应

让我们回顾第2节中描述的简单随机响应机制，用于评估尴尬或非法行为的频率。让XYZ成为这样一种活动。面对查询“你过去一周参与过XY吗？”,被告被指示执行以下步骤：

1.翻转一枚硬币。

2.如果是尾巴，那就回应如实。

3.如果是头部，则翻转第二枚硬币，如果是头部则回答“是”，如果是尾

部则回答“否”。

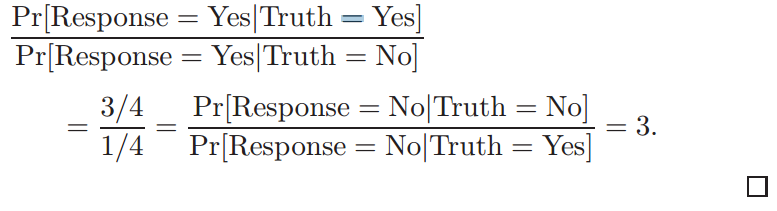
随机响应背后的直觉是它提供了“合理的可否定性”。例如，可能已经提供了“是”的响应，因为第一和第二次硬币翻转都是头部，其发生概率为1/4。换句话说，*隐私是通过过程获得的*，没有“好”或“坏”的反应。获得答复的过程会影响答复的合法解释方式。如下一个声明所示，随机响应是差分隐私的。

**声明3.5**。上述随机化响应的版本是差分隐私。

***证明。***固定受访者。案例分析表明

。

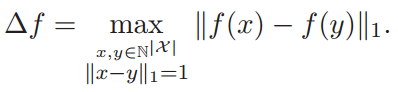
具体来说，当事实为“Yes”时，如果第一个硬币出现反面（概率1/2）或第一个和第二个出现正面（概率1/4），结果将为“Yes”，而（第一个出现头部，第二个出现尾部;概率为1/4）。将类似的推理应用于“No”答案的情况，我们得到：



3.3拉普拉斯机制

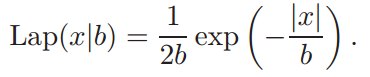
数值查询，函数，是最基本的数据库查询类型之一。这些查询将数据库映射到k个实数。决定我们回答此类查询准确程度的重要参数之一是它们的灵敏度：

定义3.1（灵敏度）。函数的灵敏度为：



函数的灵敏度捕获了单个人的数据在最坏的情况下可以改变函数的大小，因此，直观地说，我们必须引入的响应中的不确定性，以隐藏单个个体的参与。实际上，我们将形式化这种直觉：函数的敏感性给出了我们为了保护隐私对输出作出扰乱的程度上限。一种噪音分布自然会带来不同的隐私。

**定义3.2（拉普拉斯分布）。**具有尺度b的拉普拉斯分布（以0为中心）是具有以下概率密度函数的分布



该分布的方差是。我们有时会用来表示具有尺度b的拉普拉斯分布，并且有时会滥用符号并且写只是为了表示随机变量。

拉普拉斯分布是指数分布的对称形式。

我们现在将定义*拉普拉斯机制*。顾名思义，拉普拉斯机制将简单地计，并用拉普拉斯分布中得到的噪声扰动每个坐标。噪声的大小将被校准为f的灵敏度（除以）.[[3]](#footnote-3)

**定义3.3（拉普拉斯机制）**。给定任何函数，拉普拉斯机制定义为：



其中是**独立同分布**来自绘制的随机变量。

**定理3.6**。拉普拉斯机制保护差分隐私。

***证明***。

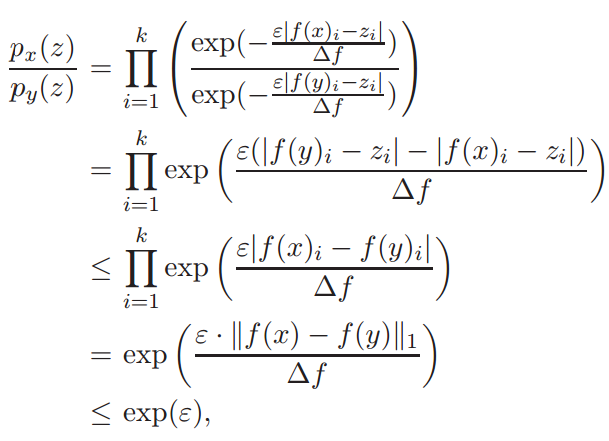
设和,，

设为函数。

设表示的概率密度函数，

设表示的概率密度函数。

我们在某个任意点比较两者，



其中第一个不等式来自三角不等式，最后一个来自灵敏度的定义和

的事实。根据对称性。

**例3.1（计数查询）**。计数查询是“数据库中有多少元素满足属性P？”形式的查询。我们将一次又一次地返回这些查询，有时以这种单纯形式，有时以分数形式（“数据库中的元素的一小部分...？“），有时使用权重（线性查询），有时采用稍微复杂的形式（例如，将应用于数据库中的每个元素并对结果求和）。计数是一个非常强大的原语。它捕获了统计查询学习模型中可学习的所有内容，以及许多标准数据挖掘任务和基本统计​​信息。由于计数查询的灵敏度为1（单个个体的添加或删除可以将计数最多改变1），因此定理3.6的直接结果是计数查询可以通过加入尺度为的噪声实现差分隐私，即通过添加从得到的噪声。预期的失真或误差为，与数据库的大小无关。

可以将固定但任意的m个计数查询列表视为向量值查询。如果没有关于该组查询的任何进一步信息，则该向量值查询的灵敏度的最坏情况界限是m，因为单个个体可能改变每个计数。在这种情况下差分隐私可以通过给每个查询的真实答案添加尺度为的噪声来实现。

我们有时会将响应大量（可能是任意的）查询的问题称为*查询发布问题*。

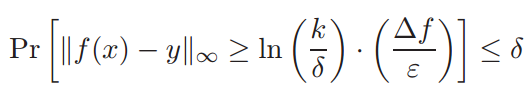
**例3.2（直方图查询）**。在特殊（但常见）的情况下，即查询在结构上是不相交的，我们可以做得更好 - 我们不一定要让噪声尺度随着查询的数量而变化。一个例子是*直方图查询*。在这种类型的查询中，宇宙被分区为单元格，查询询问每个单元格中有多少数据库元素。因为单元格是不相交的，所以添加或删除单个数据库元素会影响一个单元格中的计数，并且该单元格的差异以1为界，因此直方图查询具有灵敏度1，可以通过将取自的独立变量添加到每个单元格中的真实计数来回答。

为了理解拉普拉斯机制对一般查询的准确性，我们使用以下有用的事实：

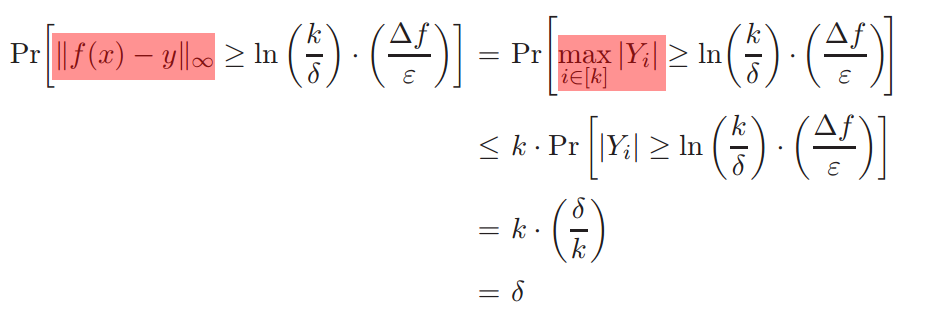
**事实3.7**。如果，则：。

这个事实与联合约束一起，给出了拉普拉斯机制准确性的简单约束：

**定理3.8。**设，并且令。然后：



**证明。**我们有(y代表啥?)：



其中倒数第二个不等式来自每个和Fact 3.7的事实。

**例3.3（名字）**。假设我们希望计算出来自10,000个潜在名字的名单中的哪些名字是2010年人口普查参与者中最常见的名字。这个问题可以表示为查询。这是一个直方图查询，因此灵敏度，因为每个人最多只能有一个名字。使用上述定理，我们看到我们可以以差分隐私同时计算所有10,000个名称的频率，并且有95％的概率，没有估计值会超过的附加误差。对于一个超过30万人的国家来说，这是一个非常低的错误！

**差分隐私选择**。示例3.3中的任务是差分隐私选择之一：结果空间是离散的，任务是产生“最佳”答案，在这种情况下是人口最多的直方图单元格。

**例3.4（最常见的医疗条件）。**假设我们希望知道在一组受访者的病史中（大约）最常见的是哪种情况，因此对于所考虑的每种情况，该组问题是该个体在这种情况下是否曾接受过该病症的诊断。由于个体可以经历许多条件，这组问题的敏感性可能很高。尽管如此，正如我们接下来所描述的，这个任务可以通过在每个计数中加入噪声来解决（注意小规模噪声，与条件总数无关）。至关重要的是，m个噪声计数本身不会被释放（尽管“获胜”计数可以在没有额外隐私成本的情况下发布）。

**Report Noisy Max。**考虑以下简单算法来确定m个计数查询中哪个具有最高值：向每个计数添加独立生成的拉普拉斯噪声并返回最大噪声计数的索引（我们忽略平局的可能性）。调用此算法报告Report Noisy Max。

请注意Report Noisy Max算法中的“信息最小化”原则：不是释放所有噪声计数并允许分析人员找到最大值及其索引，只有与最大值对应的索引才会公开。由于个体的数据可以影响所有计数，因此计数向量具有高的灵敏度，特别是，如果我们想要使用拉普拉斯机制释放所有计数，则需要更多的噪声(的？已解决)。

**声明3.9**。 Report Noisy Max算法是差分隐私。

***证明***。

给定。

令，表示数据库为时的计数向量。我们使用两个属性：

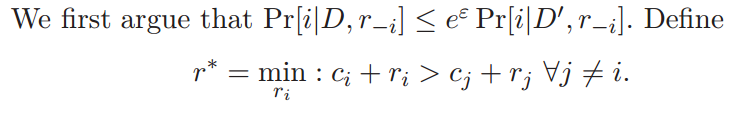
1.*计数的单调性*。对于所有;和(单调)

2. Lipschitz属性。对于所有(c相差不超过1)。

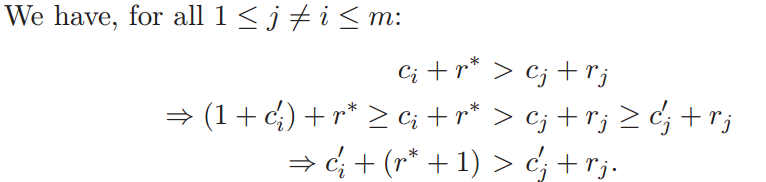
给定任何。We will bound from above and below the ratio of the probabilities that i is selected with D and with D’.。

给定r\_i(这啥啊，噪声？)，来自的抽取*(draw)*，用于除第i次计数以外的所有噪声计数。我们将独立讨论每个r-i。我们使用符号来表示报Report Noisy Max算法的输出为i的概率，以ξ为条件。

我们认为

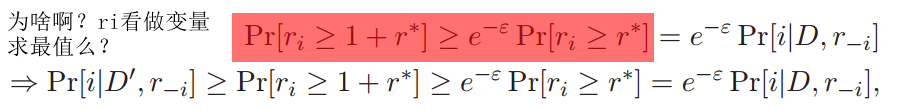


注意，固定r-i(什么意思？)时，当数据库为D时，输出将是i（argmax噪声计数）**当且仅当**时。



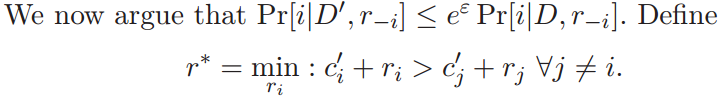
因此，**如果**，那么当数据库为时，第i个计数将是最大值并且

噪声矢量是（ri，r-i）。下面的概率关于的选择。



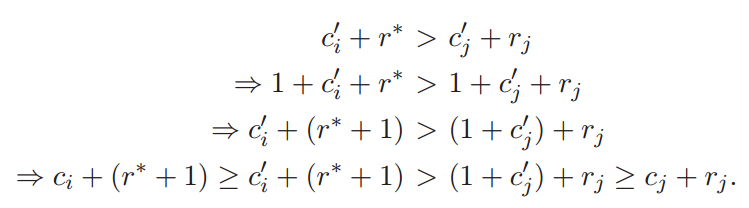
乘以后得到我们想要显示的结果：。

我们现在认为

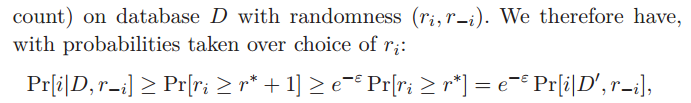


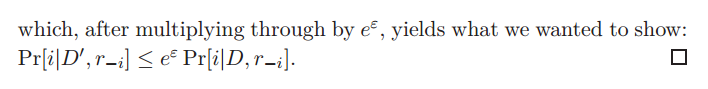
注意，有了固定的r-i，当数据库是时，输出将是i（argmax噪声计数）当且仅当。

我们有



因此，如果，则i将是具有随机性（ri，r-i）的数据库D上的输出（argmax噪声计数）。因此，





该证据很容易扩展，以显示最大噪声计数的释放，而不仅仅是其索引，不会进一步损失隐私。

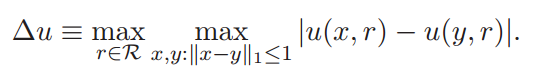
3.4指数机制

在“最常见的名称”和“最常见的条件”示例中，我们估计的响应（名称或医疗条件）的“效用”使用拉普拉斯噪声计数并报告噪声最大值。在两个例子中，响应的效用与产生的噪声值直接相关;也就是说，名称或条件的流行度以与噪声幅度相同的比例和相同的单位适当地测量。

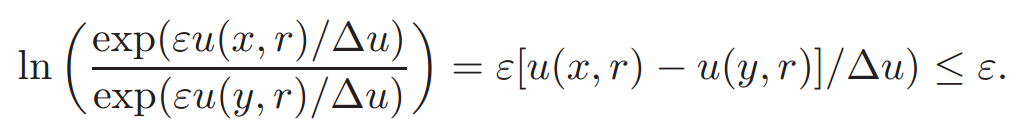
*指数机制*是为我们希望选择“最佳”响应的情况而设计的，但直接在计算数量上添加噪声可以完全破坏其值，例如在拍卖中设定价格，目标是最大化收入，以及在最优价格上添加少量正噪声（为了保护投标的隐私）可以大大减少由此产生的收入。

**实例3.5（南瓜）。**假设我们有丰富的南瓜和四个投标人：A，F，I，K，其中A，F，I各自出价1.00美元，K出价3.01美元。什么是最优价格？ 3.01美元的收入为3.01美元， 3.00美元和1.00美元收入为3.00美元，但3.02美元，收入为零！

指数机制是用于回答具有任意实用程序（和任意非数字范围）的查询的自然构建块，同时保留差分隐私。给定一些任意范围R，指数机制是根据某些效用函数定义的，将数据库/输出对映射到效用得分。直观地，对于固定数据库x，用户期望该机制输出具有最大可能效用分数的R的一些元素。请注意，当我们谈论效用分数的敏感性时，我们只关心U1对其数据库参数的敏感度；它在其范围参数中可以任意敏感：



指数机制背后的直觉是以与成正比的概率输出每个可能的r∈R，因此隐私损失约为：



这种直观的观点忽略了当数据库中的另一个人导致某些元素的效用降低而其他元素的效用增加时出现的标准化术语的一些影响。接下来定义的实际机制为标准化术语的变化保留了一半的隐私预算。

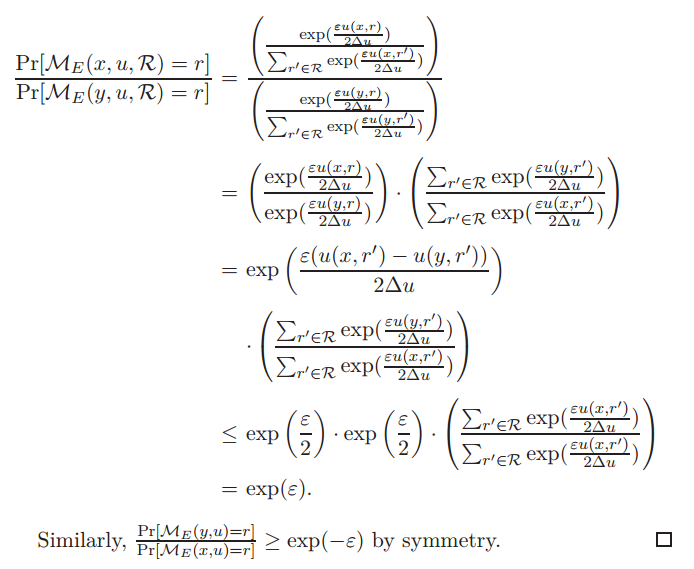
**定义3.4（指数机制）**。指数机制以与成比例的概率选择并输出元素r∈R。

指数机制可以在大的任意域上定义复杂分布，在问题的自然参数中，u的取值范围是超多项式大时，可能无法有效地实现指数机制。

回到南瓜示例，数据库x上的价格p的效用仅仅是当价格为p且需求曲线如x所述时，获得的利润。重要的是潜在价格范围与实际出价无关。否则，将存在一个数据集中具有非零权重并且邻近集合中具有零权重的价格，这违反了差分隐私。

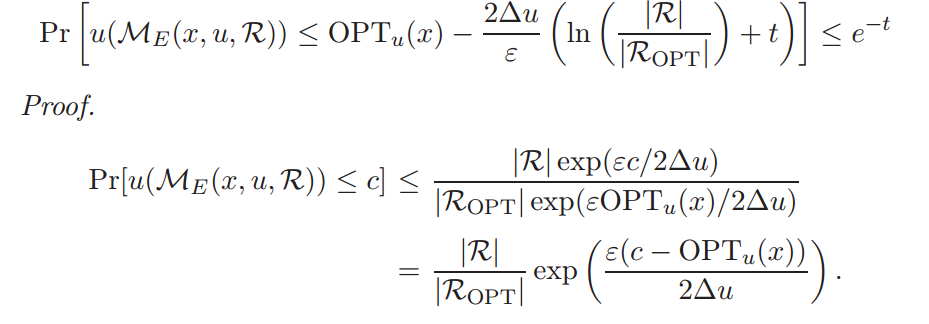
**定理3.10。**指数机制保留差分隐私。

***证明***。为清楚起见，我们假设指数机制的范围R是有限的，但这不是必需的。与所有差分隐私证明一样，我们考虑实际的指数机制在两个相邻数据库上和(即，)输出某个元素r∈R的概率的比率。



指数机制通常可以提供强大的效用保证，因为当质量得分下降时，它会以指数方式快速折扣结果。对于给定的数据库x和给定的效用度量，令表示数据库x中任何元素r∈R的最大效用得分。我们将限制指数机制返回R的“好”元素的概率，其中good将以来衡量。结果是，返回元素r的效用得分极不可能比低的加法因子。

**定理3.11。**给定数据库x，令表示R中获得效用分数的元素集。然后：(没看明白)

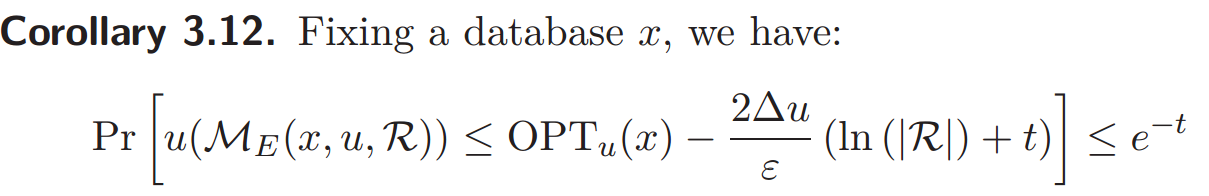


不等式来自观察，即每个的具有un-normalized概率质量最大是，因此整组这样的“坏”元素r具有总的un-normalized概率质量最大是。相反，我们知道至少存在个的元素，因此un-normalized质量，因此这是归一化(normalization)项的下限。

该定理来自插入适当的c值。

因为我们总是有，我们可以更常用的使用以下简单的推论：

**推论3.12。**给定数据库x，我们有：



如定理3.11和推论3.12的证明所示，指数机制可以特别容易分析。

**例3.6（两个中最好的）。**考虑确定两个医疗条件A和B中哪一个更常见的简单问题。对于条件A，将两个真实计数设为0，对于条件B，设置c> 0。我们的效用概念将与实际计数相关联，因此计数较大的条件具有较高的效用且。因此，A的效用为0，B的效用为c。使用指数机制，我们可以立即应用推论3.12，看到观察（错误）结果A的概率最多为。

分析Report Noisy Max似乎更复杂，因为它需要了解在（概率1/4）情况下发生的情况，当添加到A的计数的噪声为正且添加到B的计数的噪声为负时。

如果向数据集添加元素不会导致函数的值减小，则*函数在数据集中是单调的*。计数查询是单调的;通过向一系列买家提供固定价格获得的收入也是如此。

考虑*Report One-Sided Noisy Arg-Max* mechanism，把噪声加入从单调效用情况下参数 的或非单调效用情况下参数的单指数分布中中提取的每个潜在输出，并报告生成的arg-max。

使用这种算法，其隐私证明几乎与*Report One-Sided Noisy Arg-Max* mechanism的相同（但当效用非单调时失去两倍），我们立即在上面的例3.6中得到结果A在c中呈指数级比结果B更不可能被选中。

**定理3.13。** *Report One-Sided Noisy Arg-Max* mechanism，当使用参数运行时，在输出上产生与指数机制相同的分布。

请注意，与*Report Noisy Max*的情况相比，当我们使用*Report One-Sided Noisy Arg-Max* mechanism时，我们无法报告噪声效用值本身。 为了看到这一点，假设在数据库x中医疗条件B的计数是c而在相邻的y中计数是c-1，并且考虑小的区间I = [c-1，c-1 +η]，对于η< 1.然后在数据库x中，I的结果概率为零，而在数据库y上则为非零，违反差分隐私

3.5组成定理

现在我们有几个用于设计差分隐私算法的构建块，了解我们如何将它们组合起来设计更复杂的算法非常重要。为了使用这些工具，我们希望两个差分隐私算法的组合本身是差分隐私的。实际上，正如我们将要看到的那样。当然，参数和必然会降低 - 考虑使用拉普拉斯机制重复计算相同的统计量，每次缩放以给出差分隐私。机制的每个实例给出的答案的平均值最终会收敛到统计数据的真实值，因此我们无法避免重复使用时隐私保证的强度会降低。在本节中，我们给出定理,当组合差分隐私子程序时，参数和如何组成。

让我们首先从容易预热开始：我们将看到单独使用差分隐私算法和差分隐私算法，当合在一起时，是差分隐私。

**定理3.14**。

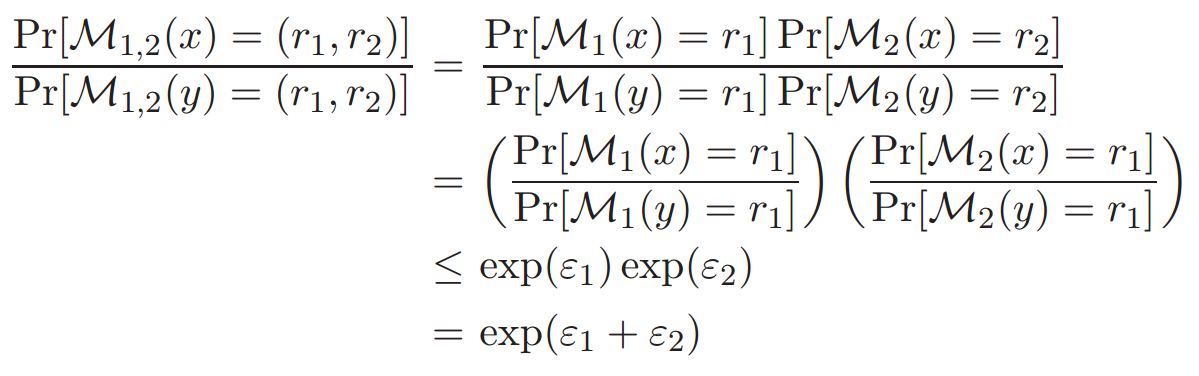
设是差分隐私算法，设是差分隐私算法。

然后他们的组合，通过映射：定义为是差分隐私的。

***证明***。

设且。

给定任何。然后：



可以重复应用组合定理以获得以下推论：

**推论3.15。**设是的差分隐私算法。然后如果被定义为，则是差分隐私。

该定理推广到差分隐私的证明见附录B：

定理3.16。设是的差分隐私算法。然后如果被定义为，则是差分隐私。

组合是“自动的”，这是一种差分隐私的优势，因为所获得的界限在没有数据库策展人的任何特殊努力的情况下保持不变。

3.5.1组成：一些技术

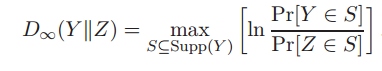
在本节的其余部分，我们将证明一个更复杂的组合定理。为此，我们需要一些定义和引理，根据分布之间的距离度量来重新描述差分隐私。在下面的小数量中，如果分母为零，那么我们将分数的值定义为无穷大（分子将始终为正）。

**定义3.5（KL-Divergence）。**从相同域获取值的两个随机变量Y和Z之间的KL-Divergence或相对熵定义为：

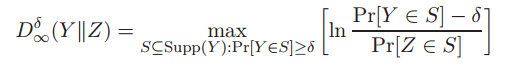


已知，当且仅当Y和Z相同地分布时才具有相等性。然而，D不是对称的，不满足三角不等式，甚至可以是无穷大的，特别是当Supp(Y)不包含在Supp(Z)中时(啥意思)。

**定义3.6（最大发散）**。取自同一个域的值两个随机变量Y和Z之间的最大散度定义为：



Y和Z之间的近似最大发散定义为：



**备注3.1**。注意机制M是

1. 差分隐私，当且仅当每两个相邻数据库x和y时，

和;并且是

1. 差分隐私, 当且仅当在每两个相邻数据库x，y：

和时。

另一个有用的距离测量是两个随机变量Y和Z之间的**统计距离**，定义为



我们说如果，则Y和Z是*close*。(绝对值)

我们将在精确最大发散和统计距离方面使用以下近似最大发散的重构：

**引理3.17。**

1. 当且仅当存在随机变量使得且 时，
2. 当且仅当存在随机变量使得,和时，我们同时具有和。

***证明。***对于第1部分，假设存在*close* ，使得。然后对于每个S，

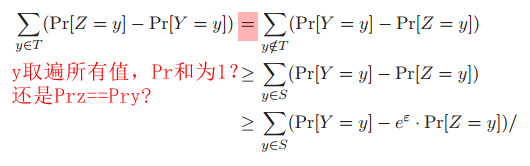


，因此。

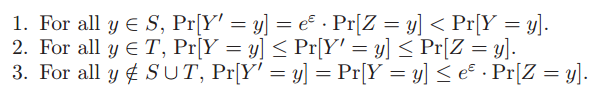
相反，假设。设。然后



此外，如果我们让，那么我们有

。

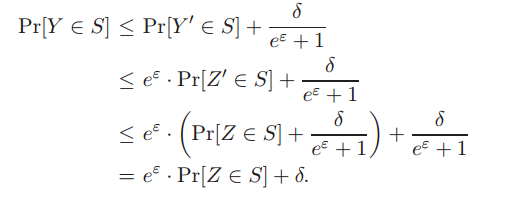
因此，我们可以通过降低S上的概率并提高T上的概率来从Y获得以满足（没看懂，为啥？）：



然后通过检查有，并且

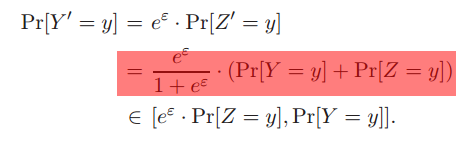


我们现在证明第2部分。假设存在如所述的随机变量和。然后，对于每组S，

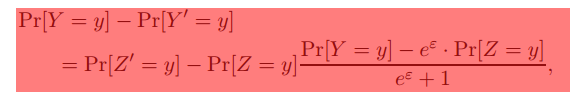


因此，并且通过对称性有。

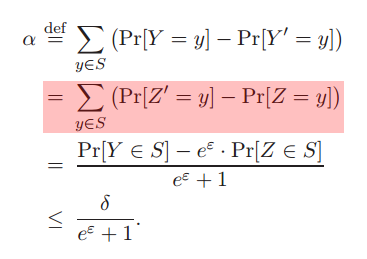
相反，给定Y和Z使得和。，我们类似于第1部分。但是，不是简单地减小S上的Y的概率质量以获得并且用消除间隙，我们也增加S上Z的概率质量。具体来说，对于每个y∈S，我们将得到



这也意味着对于y∈S，我们得到：

，

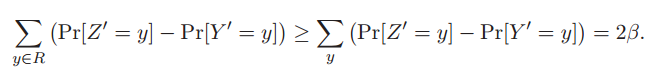
因此α



类似地，在集合上，我们可以减小Z的概率质量，并将Y的概率质量增加使得对于每个，我们得到。

如果 ，那么对于所有我们可以得到和，给定和。

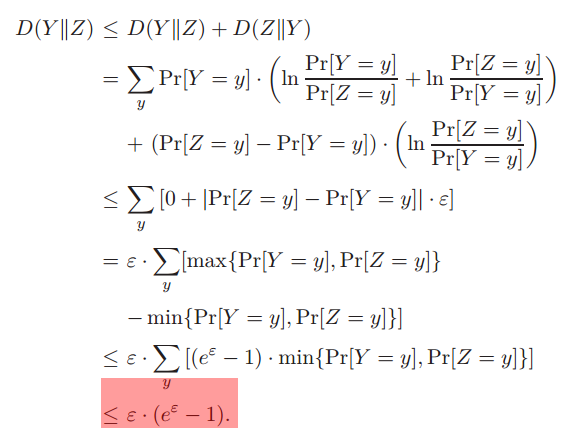
如果，比如说，那么我们还需要增加Y'的概率质量，并将S'的质量在以外的点上减去总共，确保概率总和为1.也就是说，如果我们试图采用上面定义的“质量函数”，那么对于每个y我们确实有和我们也有和。然而，这意味着如果我们让，那么

。

那么我们可以把Y的概率在R上的质量点增加β，并且Z的概率在R上的质量点减少β，同时保留了对于所有y∈R，的性质。得到的Y'和Z'具有我们想要的性质：和。

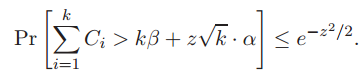
**引理3.18**。 假设随机变量Y和Z满足且。 然后.

***证明。*** 我们知道，对于任何Y和Z，情况是（通过“对数不等式”），因此它足以约束.我们得到：

。

**引理3.19（Azuma的不等式）**。 让是实值随机变量，使得对于每个，并且对于每个，我们有。

然后对于每个，我们有

。

3.5.2高级构图

除了允许参数降低得更慢之外，我们希望我们的定理能够处理更复杂的合成形式。但是，在我们开始之前，我们必须讨论组成的确切含义。我们希望我们的定义涵盖以下两个有趣的场景：

1. 在同一数据库上重复使用差分隐私算法。这允许多次重复使用相同的机

制，以及来自任意私有构建块的差分隐私算法的模块化构造。

1. 在不同数据库上重复使用差分隐私算法，但可能包含与同一个人有关的

信息。这允许我们推断单个人的累积隐私损失，其数据可能分布在多个数据集中，每个数据集可以以差别私密的方式独立使用。由于新数据库一直在创建，并且攻击者实际上可能会影响这些新数据库的构成，因此这是与反复查询单个固定数据库根本上不同的问题。

我们希望模拟组合，其中攻击者可以自适应地影响输入到未来机制的数据库，以及对这些机制的查询。设F是数据库访问机制的一族。 （例如，F可以是所有差分隐私机制的集合。）对于概率对手A，我们考虑两个实验，实验0和实验1，定义如下。

**家庭F和对手A的实验b**：

对于：

1. A输出两个相邻的数据库和，机制和参数。

2. A接收。

我们允许上面的对手A在整个实验中都是有状态的，因此它可以根据先前机制的输出自适应地选择数据库，机制和参数。我们将A的实验*视图*定义为A的硬币抛出和所有机制输出。（，和都可以从这些中重建。）

从直觉上理解，考虑一个对手总是选择来保存Bob数据和不同只在于Bob的数据被删除。然后实验0可以被认为是“现实世界”，其中Bob允许他的数据用于许多数据发布，而实验1作为“理想世界”，其中这些数据发布的结果不依赖于Bob的数据。我们对隐私的定义仍然要求这两个实验彼此“接近”，就像差分隐私定义所要求的那样。 Bob的直观保证是，在给定所有k机制的输出的情况下，对手“无法分辨”Bob的数据是否曾被使用过。

**定义3.7。**如果对于每个对手A，我们具有，其中表示上述k组成实验b中的A的视图, 我们说数据库访问机制的族F满足k-自适应组合下的差分隐私。

在k-自适应组合下的差分隐私相应的地要求。

**定理3.20（高级组合）**。对于所有，差分隐私机制的类满足k-自适应组合下差分隐私：

。

***证明。***对手A的视图由形式为的元组组成，其中r是A的硬币投掷和是机制的输出。

设。

我们将证明，因此对于每个集合S，我们有

。

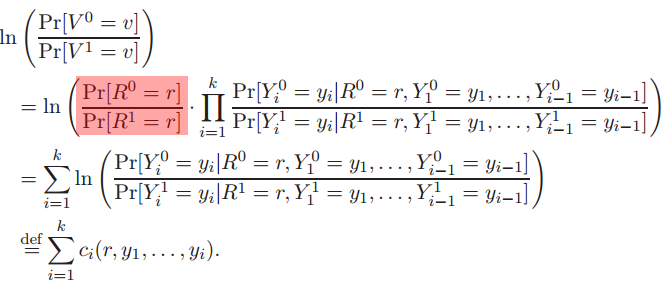
这相当于说。

剩下证明。

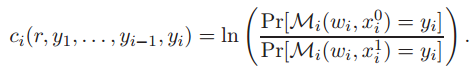
设随机变量表示实验0中A的视图，

实验1中A的视图。

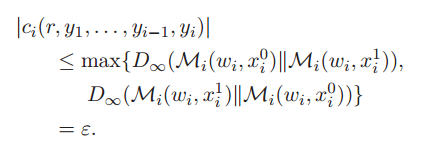
固定视图，我们有



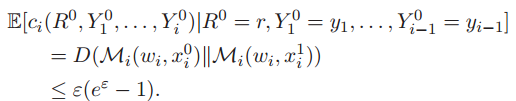
现在对于每个前缀，我们在下，分析了随机变量的期望值和最大可能值。一旦前缀被固定，还下一对数据库和，机制和由A输出的参数都确定了（在实验0和1中）。因此，根据分布。此外，对于任何值，我们有

。

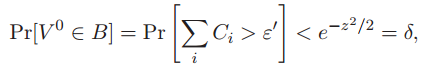
通过差分隐私，这受到的限制。我们也可以推理如下：

。

通过引理3.18，我们得到：

。

因此，我们可以将Azuma不等式应用于随机变量，其中，以推导出预期的

。

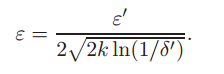
为了将证明扩展到差分隐私机制的组成，对于，我们使用引理3.17（第2部分）中的近似最大发散的特征来将分析减少到与与 不可区分的序列相同的情况。具体来说，对于由对手A选择的每个差分隐私机制和统计距离的三角不等式，使用引理3.17，第2部分，可以得出是随机变量这样对于每个前缀，如果我们以为

条件，则认为且。

这足以表明。由于是W，因此引理3.17，第1部分给出。

如果我们希望对于给定的的确保差分隐私，一个直接且有用的推论告诉我们对于k个机制中每个机制安全地选择。

**推论3.21。** 给定目标隐私参数且，为了确保k机制上的累积隐私损失，每个机制都是不同地是私有的，其中

。

***证明***。 定理3.20告诉我们所有,组合将是，其中。 当时，我们有。

请注意，上述推论给出了如何设置以在组合下获得所需隐私参数的粗略指导。当人们关心优化常数时（在处理实际实现时会这样做），通过直接应用组合定理，可以更紧密地设置。。

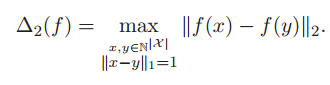
**例3.7**。假设，在他的一生中，Bob是k = 10,000差分隐私数据库的成员。假设这些数据库之间没有协调 - 任何给定数据库的管理员甚至可能都不知道其他数据库的存在 - 的值应该是什么，这样，在他的一生中，Bob的累积隐私损失受到的限制，概率至少为?。定理3.20表示，取就足以得到。假设在不同的差分隐私数据库之间没有协调，这对于任意对手来说基本上是最优的。

那么我们可以用非平凡的准确度回答多少个查询？在大小为n的数据库中，如果误差是o（n）阶，则说准确性并不重要。定理3.20表示，对于和的固定值，有可能以非平凡的精度回答接近计数的查询。类似地，人们可以回答接近n个查询，同时仍然具有噪声o（√n） - 即，噪声小于采样误差。我们将看到，通过协调添加到各个响应的噪声，可以显着改善这些结果，在某些情况下甚至可以处理具有仅略大于√n的噪声的指数数量的查询。事实证明，这种协调是必不可少的：没有协调，先进构成定理的界限几乎是紧的。

3.5.3拉普拉斯与高斯

除了添加拉普拉斯噪声，另一种方法是添加高斯噪声。在这种情况下，我们不是将噪声缩放到l1灵敏度Δf，而是缩放到灵敏度：

**定义3.8（敏感度）**。函数的灵敏度为：



具有参数b的高斯机制在k坐标中添加具有方差b的零均值高斯噪声。附录A证明了以下定理。

**定理3.22。**设是任意的。对于，具有参数的高斯机制是差分隐私。

高斯噪声的优点之一是为隐私增加的噪声与其他噪声源相同;而且，两个高斯的总和是高斯的，因此隐私机制对统计分析的影响可能更容易理解和纠正。

这两种机制在组合下产生相同的累积损失，因此即使每个单独计算的隐私保证较弱，许多计算的累积效应也是可比较的。而且，如果足够（例如，次多项式）小，实际上我们将永远不会遇到保证的弱点。

也就是说，相对于拉普拉斯噪声的经验，高斯噪声存在理论上的缺点。考虑Report Noisy Max（具有拉普拉斯噪声）, 在每个候选在数据库x上与在其邻居y上输出具有相同质量分数的情况下。与候选输出的数量无关，该机制产生差分隐私。相反，如果我们使用高斯噪声并报告最大值，并且如果候选者的数量与相比较大，那么我们将精确地选择具有大的高斯噪声的噪声 - 噪声以小于的概率发生。当我们远远高于高斯的尾部时，我们不再能保证观察值在x和y之间可能出现的因子内。

3.5.4关于组成的评论

分析组合下的累积隐私损失的能力使我们能够处理差分隐私数据库可以提供的内容。一些观察是有序的。

**弱定量**。假设攻击者总是选择来保存Bob的数据，并且是相同的数据库但删除了Bob的数据。通过适当选择参数，定理3.20告诉我们，对手 - 包括知道甚至selects(!)数据库对的对手 - 在确定b∈{0,1}的值方面几乎没有优势。这是一种固有的弱量化。我们可以确保对手不太可能将现实与任何给定的替代方案区分开来，但我们无法同时确保所有替代方案。如果有一个数十亿的数据库，但Bob只是其中10,000个的成员，那么我们就不会同时从一万亿减去一万保护Bob的缺席。这类似于差分隐私的定义中的量化，其中我们提前确定了一对相邻的数据库，并且认为这两个数据库的输出几乎同等可能。

**人与鬼。**直观地说，每个记录具有少量比特的差异差分隐私数据库比具有包含我们整个病史的选择相同的差分隐私数据库保护性差。那么在什么意义上我们的主要隐私措施，，告诉我们在存储数据的复杂性和敏感性方面存在巨大差异的数据库相同的东西？答案在于组成定理。想象一个由两种生物居住的世界：幽灵和人类。两种类型的生物都表现相同，以相同的方式与他人互动，以同样的方式写作，学习，工作，笑，爱，哭，重现，生病，恢复和年龄。唯一的区别是幽灵在数据库中没有记录，而人类有记录。隐私对手的目标是确定给定的50岁，“目标”是鬼还是人。事实上，敌人有50年的时间来这么做。对手不需要保持被动，例如，她可以组织临床试验并招募她选择的患者，她可以创建人类来填充数据库，有效地创建最坏情况（隐私）数据库，她可以将目标暴露给25岁时和35岁时的化学品，依此类推。她可以知道可能输入任何数据库的目标的所有内容。她可以知道目标所在的数据库是目标人类。组合定理告诉我们，每个数据库的隐私保证 - 无论数据类型，复杂性和灵敏度 - 都为人/鬼位提供了类似的保护。

3.6稀疏矢量技术

拉普拉斯机制可用于回答自适应选择的低灵敏度查询，我们从组成定理中知道隐私参数与回答的查询数（或其平方根）成比例地降低。不幸的是，我们经常会遇到很多问题需要回答 - 许多人使用独立的扰动技术来产生合理的隐私保证，即使是第3.5节的高级组成定理。但在某些情况下，我们只会关心了解超过某个阈值的查询的身份。在这种情况下，我们可以希望通过丢弃远低于阈值的查询的数字答案来获得朴素的分析，并仅仅报告它们确实低于阈值。 （如果我们选择的话，我们也可以获得上述阈值查询的数值，只需要很少的额外费用）。这类似于我们在3.3节中Report Noisy Max机制中所做的，并且实际上迭代该算法或指数机制将是非交互式或离线情况的选项。

在本节中，我们将展示如何在在线设置中分析此方法。该技术很简单 - 添加噪声并仅报告噪声值是否超过阈值 - 我们的重点是分析，表明隐私只会随着实际位于阈值以上的查询数量而降低，而不是查询总数。如果我们知道位于阈值之上的查询集远小于查询总数（即，答案向量是稀疏的），则可以节省大量成本。

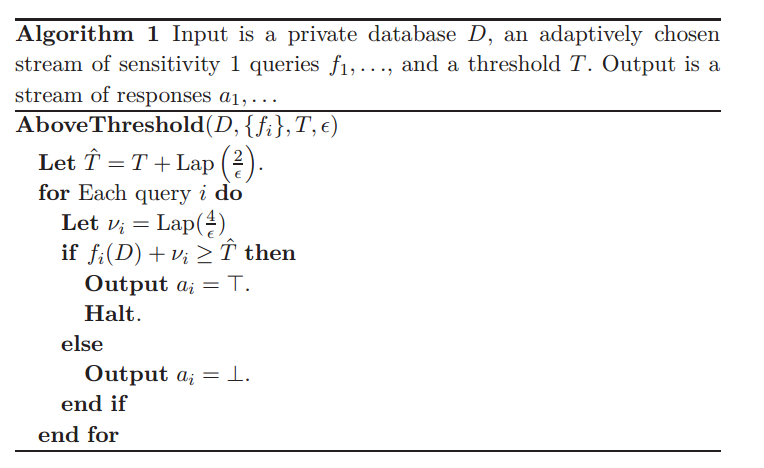
更详细一点，我们将考虑一系列事件 - 每个查询一个 - 如果在数据库上评估的查询超过给定（已知，公共）阈值，则会发生这一事件。我们的目标是发布一个位向量，指示每个事件是否已经发生。当呈现每个查询时，该机制将计算噪声响应，将其与（公知的）阈值进行比较，并且如果超过阈值，则揭示该事实。由于隐私证明中的技术原因（定理3.24），该算法使用阈值T的噪声版本.虽然T是公共的，但是噪声版本不是。

以下分析不会导致每个可能的查询的隐私丢失，而是仅针对接近或高于阈值的查询值导致隐私成本。

**设定**。令m表示灵敏度1查询的总数，其可以自适应地选择。在不失一般性的情况下，预先确定了单个阈值T（或者，每个查询可以具有其自己的阈值，但结果不变）。我们将向查询值添加噪声并将结果与​​T进行比较。正结果意味着有噪声的查询值超过阈值。我们期望少量c的噪声值超过阈值，并且我们仅释放高于阈值的噪声值。该算法将在其停止条件下使用c。

我们将首先分析在c = 1高于阈值的查询之后算法停止的情况，并且无论查询的总序列有多长，都表明该算法是差分隐私的。然后，我们将使用我们的组合定理分析的情况，并导出和 差分隐私的界限。

我们首先讨论，AboveThreshold，专门针对只有一个阈值以上查询的情况的算法，是私有且准确的。



**定理3.23**。 AboveThreshold是差分隐私。

***证明***。给定任意两个相邻的数据库D和D'。

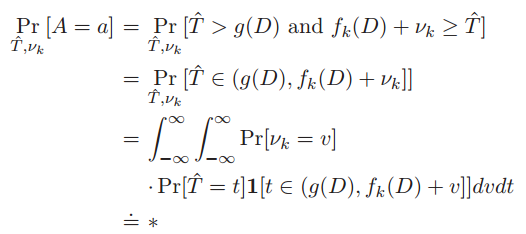
设A表示表示输出的随机变量，

设A'表示表示输出的随机变量。

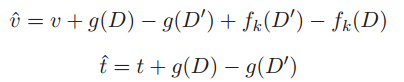
算法的输出是这些随机变量的一些实现，，并且具有对于所有和的形式。算法内部有两种类型的随机变量：噪声阈值和对k个查询的扰动，。对于以下分析，我们将给定的（任意）值并获取和随机性的概率。定义以下数量，表示任何查询在D上评估的最大噪声值：

。

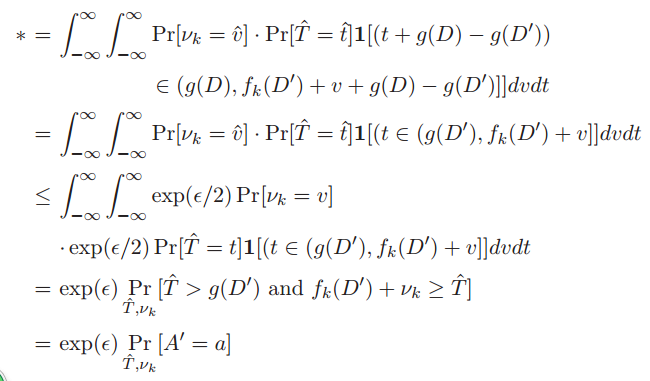
在下文中，我们将滥用符号并将写为在（类似于）处评估的的概率密度函数的简写，并写以表示事件x的指示符函数。注意给定的值（使g（D）成为确定性量），我们得到：

。

我们现在改变变量。定义：

，

并注意任何和.这是因为每个查询是1敏感的，因此数量也是1敏感的。应用这种变量变化，我们得到：



，其中不等式来自我们对和的界限以及拉普拉斯分布的概率密度函数格式。

**定义3.9（准确性）**。如果除了概率最大β之外，算法不会在之前停止，并且对于所有并且对于所有，我们将说出一个输出答案流以响应k个查询的流的算法相对于阈值T是准确的。

算法1中可能出现什么问题？噪声阈值可以离T很远，比如说。此外，小计数可以添加如此多的噪声，因此报告为高于阈值（即使阈值接近正确），并且大计数可以报告为低于阈值。所有这些都发生在概率指数小的中。总之，我们可能在选择噪声阈值时遇到问题，或者我们可能会遇到一个或多个单独噪声值的问题。当然，我们可能这有两种错误，所以在下面的分析中我们为每种类型分配的概率。

**定理3.24**。对于任何k个查询序列使（即接近高于阈值的唯一查询可能是最后一个），是精确的：。

**证明。**如果我们能证明除概率最大为外，这个定理将被证明。

如果是这种情况，那么对于任何，我们有：

，

或换句话说：。

同样，对于任何，我们有：。

对于任何，我们也将如此：，意味着。因此，在回答k个查询之前，算法不会停止。

我们现在完成证明。

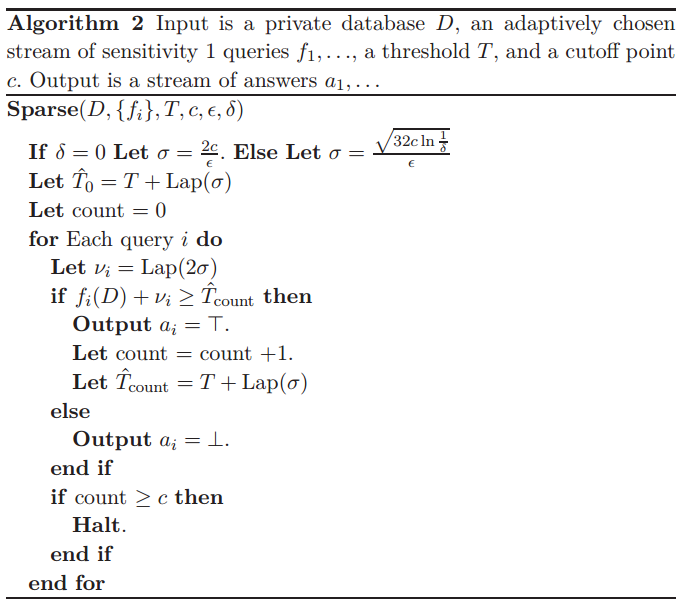
回想一下，如果，那么：。因此我们有：。

设置这个​​数量最多为，我们发现我们需要。

同样，通过联合约束，我们有：。将此数量设置最多为，我们发现我们需要。这两个主张相结合证明了这个定理。

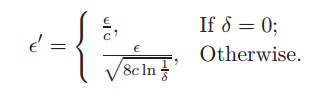
我们现在展示如何使用组合处理多个“超过阈值”的查询。

稀疏算法可以被认为如下：当查询进入时，它会重复调用AboveThreshold。每次报告超过阈值的查询时，该算法只会在AboveThreshold的新实例化上重新启动剩余的查询流。它在重新启动AboveThreshold c次之后停止（即在出现c次高于阈值查询之后）。 AboveThreshold的每个实例都是private，因此组合定理适用。



**定理3.25。**稀疏是 差分隐私的。

***证明。***我们观察到Sparse完全等同于以下过程：我们使用AboveThreshold提供的答案在我们的查询流上运行设置

。

当AboveThreshold停止时（在1次高于阈值查询之后），我们只需在剩余的流上重新启动Sparse ，并以此方式继续，直到我们重新启动AboveThreshold c次。在AboveThreshold 的第c次重新启动后，我们也停止了。我们已经证明是差分私有的。最后，通过高级组合定理（定理3.20），差分私有算法的C应用是差别私有，并且差分私有算法的c应用是private。

通过再次观察Sparse仅包含对AboveThreshold的c次调用，仍然可以证明Sparse的准确性。我们注意到，如果对AboveThreshold的这些调用中的每一个都是准确的，那么Sparse是 准确的。

**定理3.26。**对于任何k个查询序列使得，如果，Sparse是精确的：

。

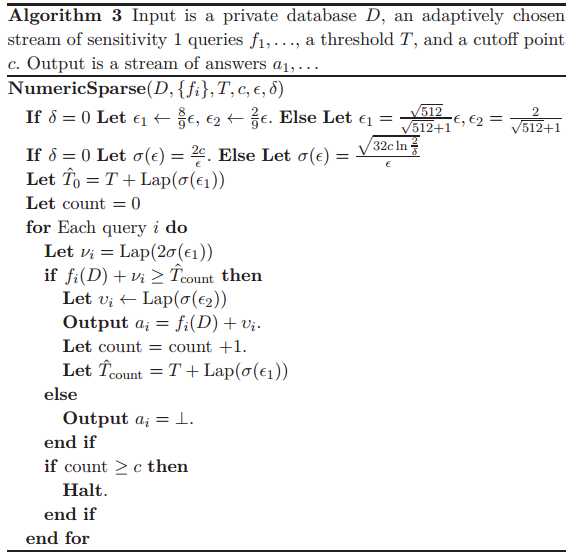
如果，稀疏是精确的：

。

***证明。***我们简单地将定理3.24设置设为，

设为 和，分别取决于或。

最后，我们给出了一个Sparse版本，它实际上输出了上述阈值查询的数值，我们可以只用常量因子精度损失。我们称这个算法为NumericSparse，它只是Sparse与Laplace机制的组合。它不输出向量，而是输出向量。



我们观察到NumericSparse是私有的：

**定理3.27。** NumericSparse是差分隐私。

***证明。***观察到

如果，则NumericSparse只是Sparse的自适应组合，

以及具有隐私参数的拉普拉斯机制。

如果，则NumericSparse是 Sparse与

具有隐私参数的拉普拉斯机制一起的组合。

因此，NumericSparse的隐私来自简单的组合。

为了讨论准确性，我们必须定义我们所说的输出流以响应一系列数值查询的机制的准确性：

**定义3.10（数字精度）。**我们将说出一个输出答案流的以响应k个查询的流的算法相对于阈值T是准确的，如果除非概率最大为β，算法在之前不停止，并且对于所有且对于所有

**定理3.28。**对于任何k个查询序列使得，如果，Sparse是精确的：

。

如果，Sparse是精确的：

。

***证明。***准确性需要两个条件：首先，对于所有。这通过稀疏的精度定理以概率保持。接下来，对于所有，它需要。这通过拉普拉斯机制的精度以概率保持。

我们到底表达(证明)了什么？如果给出一系列查询，保证其中只有c个答案高于，我们可以回答那些高于给定阈值T的查询，直到误差α。如果如果我们事先知道这些大于阈值的大型查询的身份，并使用Laplace机制来回答它们，那么在同样的隐私保证下，这个精度与我们所能得到的精度相等，最高可达常数和一个因子。也就是说，稀疏向量技术允许我们几乎“免费”地找出这些大型查询的身份，只对不相关的查询支付对数级的费用。通过尝试使用指数机制查找大型查询然后使用拉普拉斯机制回答它们，我们可以获得同样的保证。然而，这种算法运行起来微不足道，而且至关重要的是，它允许我们自适应地选择我们的查询。

3.7书目笔记

随机响应是华纳[84]（差分隐私超过四十年！）。拉普拉斯机制归功于Dwork等人 [23]。指数机制由McSherry和Talwar [60]发明。在[21]中声称定理3.16（简单组成）;附录B中出现的证明是由于Dwork和Lei [22]; McSherry和Mironov获得了类似的证明。第3.5.1节和第3.5.2节中的材料几乎逐字逐句地来自Dwork等人[32]。[32]之前的组合是非正式建模的，就像我们对简单组合边界所做的那样。对于应用于单个数据库的特定机制，由于Dinur，Dwork和Nissim [18,31]（差分隐私的定义之前）存在“信心演变”，表明k-组合中的隐私参数，如果我们愿意容忍δ（可忽略不计）的损失（对于），折叠成分只需要像一样恶化。定理3.20将这些参数推广到任意差分私有机制，

声称如果没有噪声协调，组成定理中的界限几乎是紧张的，这是由于Dwork，Naor和Vadhan [29]。稀疏矢量技术是由Dwork，Naor，Reingold，Rothblum和Vadhan [28]（引理4.4的证明中的指示矢量）引入的技术的抽象。它随后被广泛使用（例如Roth和Roughgarden [74]，Dwork，Naor，Pitassi和Rothblum [26]，以及Hardt和Rothblum [44]）。在我们对该技术的介绍中，定理3.23的证明归功于Salil Vadhan。

1. . 第8.1节中证明了这一结果，适用于保护隐私数据的所有技术，而不仅仅是差异隐私。 [↑](#footnote-ref-1)
2. . 然而，随着群体规模的扩大，隐私权保障也在恶化，这正是我们想要的：很明显，如果我们用完全不同的回答群体（例如健康的青少年）取代整个被调查人群（例如癌症患者），我们应该得到不同的答案。经常每天跑三英里的受访者比例。虽然（ε，δ）-差异隐私也有相似之处，但近似项δ受到了很大影响，我们只获得（kε，ke（k−1）εδ）-大小为k的群体的差异隐私。 [↑](#footnote-ref-2)
3. . 或者，使用校正为Δfln（1 /δ）/ε的方差的高斯噪声，可以实现（ε，δ） - 差异隐私（见附录A）。 拉普拉斯机制的使用更清晰，两种机制在组成下表现相似（定理3.20） [↑](#footnote-ref-3)