Privacy, Data Privacy, and Differential Privacy

James Bailie

(Harvard University)

16/07/2024, 11.00am

Department of Statistics, Ludwigstr. 33, Room 144
and online via Zoom (Link)
(Meeting-ID: 683 0699 4223; Password: StatsCol23)

This talk beckons inquisitive audiences to explore the intricacies of data privacy. We journey back to the late 19th century, when the concept of privacy crystallised as a legal right. This change was spurred by the vexations of a socialite's husband, harried by tabloids during the emergence of yellow journalism and film photography. In today's era, marked by the rise of digital technologies, data science, and generative AI, data privacy has surged to become a major concern for nearly every organisation. Differential privacy (DP), rooted in cryptography, epitomises a significant advancement in balancing data privacy with data utility. Yet, as DP garners attention, it unveils complex challenges and misconceptions that confound even seasoned experts. Through a statistical lens, we examine these nuances. Central to our discussion is DP's commitment to curbing the relative risk of individual data disclosure, unperturbed by an adversary's prior knowledge, via the premise that posterior-to-prior ratios are constrained by extreme likelihood ratios. A stumbling block surfaces when 'individual privacy' is delineated by counterfactually manipulating static individual data values, without considering their interdependencies. Alarmingly, this static viewpoint, flagged for its shortcomings for over a decade (Kifer and Machanavajjhala, 2011, ACM; Tschantz, Sen, and Datta, 2022, IEEE), continues to overshadow DP narratives, leading to the erroneous but widespread belief that DP is impervious to adversaries' prior knowledge. Turning to Warner's (1965, JASA) randomised response mechanism—the first recorded instance of a DP mechanism—we show how DP's mathematical assurances can crumble to an arbitrary degree when adversaries grasp the interplay among individuals. Drawing a parallel, it's akin to the folly of solely quarantining symptomatic individuals to thwart an airborne disease's spread. Thus, embracing a statistical perspective on data, seeing them as accidental manifestations of underlying essential information constructs, is as vital for bolstering data privacy as it is for rigorous data analysis. Finally, unifying the many types of DP as different kinds of Lipschitz continuity on the data-release mechanism (hence the 'differential' in differential privacy), we elicit from existing literature five necessary building blocks for a DP specification.

They are, in order of mathematical prerequisite, the protection domain (data space), the scope of protection (data multiverse), the protection unit (unit for data perturbation), the standard of protection (measure for output variations), and the intensity of protection (privacy loss budget). In simple terms, these are respectively the "what", "where", "who", "how", and "how much" questions of DP. We answer these questions for data swapping – a traditional statistical disclosure control method used, for example, in the 1990, 2000 and 2010 US Decennial Censuses – drawing parallels with the recent implementation of DP in their 2020 Census and unveiling the nuances and potential pitfalls in employing DP as a theoretical yardstick for privacy methodologies.

Biography:
As of September 2020, James Bailie is a PhD student in statistics at Harvard, advised by Prof. Xiao-Li Meng and with the support of a Fulbright Future Scholarship.
His primary research interest is currently statistical privacy – in particular differential privacy, other formal privacy methods and disclosure limitation more generally. He is also a member of the AI and Global Development Lab. Previously, he was a researcher in the Data Access and Confidentiality Methods Unit at the Australian Bureau of Statistics. His research has been recognised through the Ken Foreman award and 2020 International Association for Official Statistics Young Statistician Prize. In 2020, he was also on secondment to the Covid-19 Taskforce at the Australian Department of the Prime Minister and Cabinet.