Privacy, Data Privacy, and Differential Privacy

A seminar by Xiao-Li Meng

Department of Statistics - Harvard University (USA)

Friday 17 Nov 2023 | 12.30 p.m.
Room Benvenuti
Department of Statistical Sciences

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 crystallized 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 organization. Differential privacy (DP), rooted in cryptography, epitomizes 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) randomized response mechanism—the pioneering 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. (This presentation is based on joint work with James Bailie and Ruobin Gong).

Università degli Studi di Padova

Dipartimento di Scienze Statistiche