Discussion of Strack & Yang, 'Privacy-preserving signals'

Ludvig Sinander
University of Oxford

Theory @ Penn State
8 December 2023

In a nutshell

Question: given random variables $\boldsymbol{\omega}$ and $\boldsymbol{\theta} \equiv f(\boldsymbol{\omega})$
what information can be conveyed about $\boldsymbol{\omega}$
without conveying any information about $\boldsymbol{\theta}$?

Answer: information about $(\boldsymbol{\omega}|\boldsymbol{\theta})$-quantiles.

Interpretations: privacy, avoiding 'disparate impact', ...

More abstractly: how & to what extent
can info be 'orthogonalised' / 'factorised'?

(important in e.g. dynamic mech design)

Setting

Probability space $(\Omega, \mathcal{F}, \mathbf{P})$ standard Borel

- random variable denoted $\boldsymbol{\omega}$ (typical realisation $\omega \in \Omega$)
  formally $\boldsymbol{\omega} : \Omega \to \Omega$ given by $\boldsymbol{\omega}(\omega) = \omega \quad \forall \omega \in \Omega$
- captures all 'fundamental' uncertainty
  (generally multi-dimensional)
- interpret as cross-sectional heterogeneity

Collection $\mathcal{P} \subseteq \mathcal{F}$ called 'privacy sets'

- interpret each $P \in \mathcal{P}$ as yes/no question ('Swedish?')
  answer = 'yes' if $\boldsymbol{\omega} \in P$, = 'no' otherwise
- non-binary questions ('Swedish, Danish or other?')
  captured by collections of binary questions
  (e.g. 'Swedish or not?' & 'Danish or not?')
- wlog assume $\mathcal{P}$ a σ-algebra

Signals

Signal: random variable $\boldsymbol{s}$ (typical realisation s)

defined on (rich) extended probability space

$(\ \Omega \times \Omega',\ \mathcal{F} \times \mathcal{F}',\ \Pr\ )$

(Authors describe by Blackwell experiment (S, π).)

Signals convey info about $\boldsymbol{\omega}$

(posterior $\Pr(\boldsymbol{\omega} \in E | \boldsymbol{s} = s)$ generally varies with s)

Signal $\boldsymbol{s}$ is privacy-preserving (PP) iff

$\boldsymbol{s}$ measurable w.r.t. $\mathcal{P} \times \mathcal{F}'$

$\iff$ $\Pr(P | \boldsymbol{s} = s) = \mathbf{P}(P) \quad \forall s,\ \forall P \in \mathcal{P}$

$\iff$ conveys no info about the questions $\mathcal{P}$.

Equivalent approach

Let f be 'question-answering function' for $\mathcal{P}$:
$\forall\omega,\quad f(\omega)$ is list of answers (yes/no) to each question in $\mathcal{P}$

(for measure-theoretic niceties [actually very simple], see Prop 1)

Define random variable $\boldsymbol{\theta} := f(\boldsymbol{\omega}) \quad \forall\omega \in \Omega$

Evidently $\boldsymbol{s}$ PP iff independent of $\boldsymbol{\theta}$

$$\iff \quad \Pr(\boldsymbol{\theta} \in T | \boldsymbol{s} = s) = \mathbf{P}(\boldsymbol{\theta} \in T) \quad \forall s, \ \forall \text{ meas'ble } T$$

$\iff$ conveys no info about the questions $\mathcal{P}$.

Can go the other way, too: if start with f,
let $\mathcal{P} := \sigma(\boldsymbol{\theta})$ (generated σ-algebra). Approaches equivalent.

Main interpretation

$\boldsymbol{\omega} = (\boldsymbol{\eta}, \boldsymbol{\theta})$ is vector of characteristics.

$\boldsymbol{\theta}$ are protected or private characteristics.

Garbling preserves PP

$$\boldsymbol{s} \text{ garbling of } \boldsymbol{s'} \quad \& \quad \boldsymbol{s'} \text{ PP} \quad \Longrightarrow \quad \boldsymbol{s} \text{ PP.}$$

Simplifying assumptions

For simplicity, assume

– $\Omega \subseteq \mathbf{R}$

– condition'l CDF $\omega \mapsto F(\omega|\theta) := \Pr(\boldsymbol{\omega} \leq \omega|\boldsymbol{\theta} = \theta)$

is <u>continuous</u> $\forall\theta$

Former is 'wlog', but $F(\cdot|\boldsymbol{\theta})$ hard to interpret in general.

The (conditional) quantile signal

Conditional quantile: $\boldsymbol{q} := F(\boldsymbol{\omega}|\boldsymbol{\theta})$. $(\boldsymbol{q}|\boldsymbol{\theta} = \theta) \sim U([0,1]) \quad \forall\theta$.

<u>'(Conditional) quantile signal'</u>: $\boldsymbol{s} = \boldsymbol{q}$.

– 'applicant is in $\boldsymbol{q}$th quantile of <u>her group</u>'

('her group' = $\boldsymbol{\theta}$, but that's kept secret)

– dist'n $(\boldsymbol{s}|\boldsymbol{\theta} = \theta)$ doesn't vary with θ $\implies$ $\boldsymbol{s}$ PP.

PP signal 2: $\boldsymbol{s} = \begin{cases} \text{'below median of her group'} & \text{if } \boldsymbol{q} \leq 1/2 \\ \text{'above median of her group'} & \text{if } \boldsymbol{q} > 1/2 \end{cases}$

– garbling of $\boldsymbol{q}$ $\implies$ $\boldsymbol{s}$ PP.

Distributions of posterior means

What matters about a signal is induced random posterior belief.

In many applications, only mean of posterior belief matters.

Random posterior mean induced by signal $\boldsymbol{s}$: $\boldsymbol{\mu} = \mathbf{E}(\boldsymbol{\omega}|\boldsymbol{s})$.

Well-known: for a CDF G, the following are equivalent:

- $\mathbf{E}(\boldsymbol{\omega}|\boldsymbol{s}) \sim G$ for some signal $\boldsymbol{s}$
- $G \leq_{\text{cvx}} F$

where F is CDF of $\boldsymbol{\omega}$ $\quad F(\omega) := \mathbf{P}(\boldsymbol{\omega} \leq \omega)$.

Theorem 2. Under simplifying assumptions,
for a CDF G, the following are equivalent:

- $\mathbf{E}(\boldsymbol{\omega}|\boldsymbol{s}) \sim G$ for some PP signal $\boldsymbol{s}$
- $G \leq_{\text{cvx}} \bar{F}$

where $\bar{F}$ is CDF of $\bar{\boldsymbol{\mu}} = \mathbf{E}(\boldsymbol{\omega}|\boldsymbol{q})$ $\quad \bar{F}(\mu) := \mathbf{P}(\bar{\boldsymbol{\mu}} \leq \mu)$.

Corollary: factorisation

Given $\mu \in \mathbf{R}$, let $\mathcal{D}_\mu = \{\text{CDFs with mean } \mu\}$.

Fact: $\mathcal{D}_\mu$ ordered by $\leq_{\text{cvx}}$ is a lattice.

Proof: for any $G \in \mathcal{D}_\mu$, write $C_G(\omega) \coloneqq \int_0^\omega G \quad \forall \omega \in \Omega$.

Well-known: $G \leq_{\text{cvx}} H$ iff $C_G \leq C_H$ pointwise.

Well-known: $\{\text{functions}\}$ ordered by 'pointwise inequality' is a lattice $\begin{pmatrix} \wedge = \text{pointwise minimum,} \\ \vee = \text{pointwise maximum} \end{pmatrix}$. ■

Corollary: factorisation

Given $\mu \in \mathbf{R}$, let $\mathcal{D}_\mu = \{\text{CDFs with mean } \mu\}$.

Fact: $\mathcal{D}_\mu$ ordered by $\leq_{\text{cvx}}$ is a lattice.

Fact + Th'm 2: simple factorisation of posterior-mean dist'ns into PP & privacy-violating components.

Namely: for any CDF G that is feasible $(G \leq_{\text{cvx}} F)$,

- 'PP component': $G \wedge_{\text{cvx}} \bar{F}$, the most informative/dispersed posterior-mean dist'n that is less informative/dispersed than G & induced by a <u>PP</u> signal.
- 'privacy-violating component': remaining variation in G.

More PP signals

Conditional quantile: $\boldsymbol{q} := F(\boldsymbol{\omega}|\boldsymbol{\theta})$. $(\boldsymbol{q}|\boldsymbol{\theta} = \theta) \sim U([0,1]) \quad \forall\theta$.

PP signal 3: $(\boldsymbol{s}|\boldsymbol{q} = q) \sim U\left(\left\{\frac{q}{n}, \frac{q}{n} + \frac{1}{n}, \ldots, \frac{q}{n} + \frac{n-1}{n}\right\}\right)$

- can recover $\boldsymbol{q}$ from $\boldsymbol{s}$: $\boldsymbol{q} = n\boldsymbol{s} \mod 1$

  $\implies$ $\boldsymbol{s}$ Blackwell-equiv. to $\boldsymbol{q}$ $\implies$ $\boldsymbol{s}$ PP
- $(\boldsymbol{s}|\boldsymbol{\theta} = \theta) \sim U([0,1]) \quad \forall\theta$

PP signal 4: $(\boldsymbol{s}|\boldsymbol{q} = q) \sim U\left(\Phi^{-1}(q)\right)$, where $\Phi : [0,1] \to [0,1]$

- special cases: $\Phi(s) = ns \mod 1$, $\Phi(s) = 1 - ns \mod 1$
- can recover $\boldsymbol{q}$ from $\boldsymbol{s}$: $\boldsymbol{q} = \Phi(\boldsymbol{s})$ $\implies$ $\boldsymbol{s}$ PP.
- normalisation: Φ measure-preserving $\left(\begin{array}{r} \boldsymbol{u} \sim U([0,1]) \\ \implies \Phi(\boldsymbol{u}) \sim U([0,1]) \end{array}\right)$

  $\implies$ $(\boldsymbol{s}|\boldsymbol{\theta} = \theta) \sim U([0,1]) \quad \forall\theta$

More PP signals

Conditional quantile: $\boldsymbol{q} := F(\boldsymbol{\omega}|\boldsymbol{\theta})$. $(\boldsymbol{q}|\boldsymbol{\theta} = \theta) \sim U([0,1]) \quad \forall \theta$.

PP signal 5: $(\boldsymbol{s}|\boldsymbol{q} = q, \boldsymbol{\theta} = \theta) \sim U\Big(\Phi_{\theta}^{-1}(q)\Big)$,

where $\Phi_\theta : [0,1] \to [0,1]$ measure-preserving

- name: <u>'reordered quantile signal (RQS)'</u>
- special case: $\Phi_\theta(s) = n(\theta)s \mod 1$
- could recover $\boldsymbol{q}$ from $\boldsymbol{s}$ and $\boldsymbol{\theta}$: $\boldsymbol{q} = \Phi_{\boldsymbol{\theta}}(\boldsymbol{s})$
- Φ_θ measure-preserving $\Longrightarrow$ $(\boldsymbol{s}|\boldsymbol{\theta} = \theta) \sim U([0,1]) \quad \forall\theta$

$\Longrightarrow$ $\boldsymbol{s}$ PP.

Characterisation of PP signals

Theorem 1. Under simplifying assumptions
($\Omega \subseteq \mathbf{R}$, $\quad \omega \mapsto F(\omega|\theta)$ continuous),

– Every PP signal is a garbling of some RQS.

– RQSs are maximally informative among PP signals.

Not directly in terms of beliefs, but can re-state it that way.

Much more 'wrinkly' than Th'm 2, I find.

A (resolved) puzzle

RQS: $(\boldsymbol{s}|\boldsymbol{q} = q, \boldsymbol{\theta} = \theta) \sim U\Big(\Phi_\theta^{-1}(q)\Big)$,

where $\Phi_\theta : [0,1] \to [0,1]$ measure-preserving.

Could recover $\boldsymbol{q}$ from $\boldsymbol{s}$ and $\boldsymbol{\theta}$: $\boldsymbol{q} = \Phi_{\boldsymbol{\theta}}(\boldsymbol{s})$.

$\overset{?}{\Longrightarrow}$ $\boldsymbol{s}$ a garbling of $\boldsymbol{q}$? (Contradicts Th'm 1!)

Yes if $\boldsymbol{\theta}$ is noise (independent of $\boldsymbol{\omega}$ & non-degenerate)

...but that's ruled out: $\boldsymbol{\theta} = f(\boldsymbol{\omega})$.

More general setting: arbitrary RVs $\boldsymbol{\omega}$ & $\boldsymbol{\theta}$.

- Th'm 1 false as stated. What replaces it?
- Conjecture: Th'm 2 remains true as stated.