Rank tests and confidence sets for matrix completion

Alexandra Carpentier[1]
Joint work with : O. Klopp, M. Löffler, and R. Nickl

Institut für Mathematik, Universität Potsdam

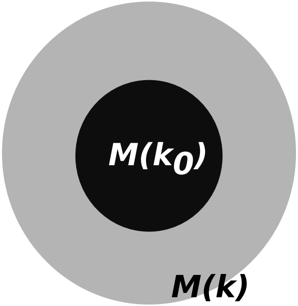
Warwick CRiSM Workshop, September 16th, 2016

[1]Supported by the DFG Emmy Noether Grant CA 1488/1-1

Testing and confidence sets

It is customary to link uncertainty quantification and decision theory.

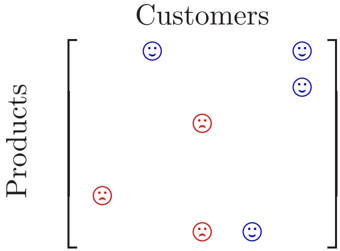
In large scale situation, this implies delicate situations and strange results, in particular when it comes to adaptive inference.



Matrix completion

Let θ be a matrix of dimension d^2. Given a small number n of noisy entries of θ, inference on θ?

Example of application : inference over a customer database.



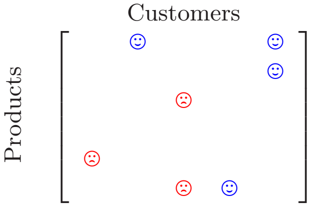
The models : with or without repetitions

Bernoulli Model

Data :

$$Y_{i,j} = (\theta_{i,j} + \varepsilon_{i,j}) B_{i,j}, \ (i,j) \in \{1, \ldots, d\}^2,$$

where $B_{i,j} \sim_{iid} \mathcal{B}(n/d^2)$ and ε is an independent noise such that $|\varepsilon| \leq 1$.



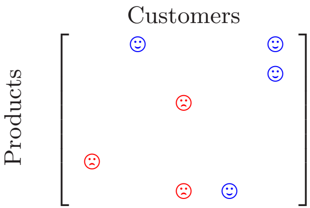
The models : with or without repetitions

Bernoulli Model

Data :

$$Y_{i,j} = (\theta_{i,j} + \varepsilon_{i,j}) B_{i,j}, \ \ (i,j) \in \{1, \ldots, d\}^2,$$

where $B_{i,j} \sim_{iid} \mathcal{B}(n/d^2)$ and ε is an independent noise such that $|\varepsilon| \leq 1$.

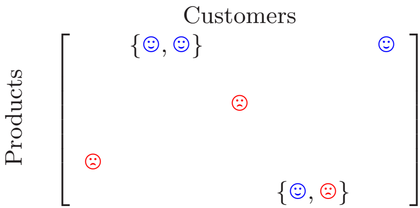


Trace Regression Model

Data :

$$Y_i = \theta_{U_i, V_i} + \varepsilon_i, \ \ i = 1, \ldots, n,$$

where $(U_i, V_i) \sim_{iid} \mathcal{U}_{\{1,\ldots,d\}^2}$ and ε is an independent noise s. t. $|\varepsilon| \leq 1$.



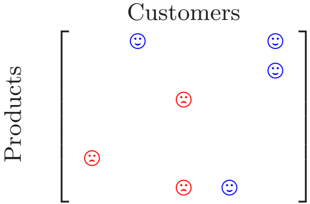
The models : with or without repetitions

Bernoulli Model

Data :

$$Y_{i,j} = (\theta_{i,j} + \varepsilon_{i,j}) B_{i,j}, \ \ (i,j) \in \{1, \dots, d\}^2,$$

where $B_{i,j} \sim_{iid} \mathcal{B}(n/d^2)$ and ε is an independent noise such that $|\varepsilon| \leq 1$.

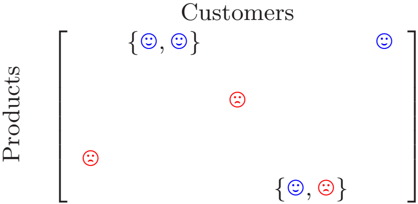


Trace Regression Model

Data :

$$Y_i = \theta_{U_i, V_i} + \varepsilon_i, \ \ i = 1, \dots, n,$$

where $(U_i, V_i) \sim_{iid} \mathcal{U}_{\{1,\dots,d\}^2}$ and ε is an independent noise s. t. $|\varepsilon| \leq 1$.



High dimensional regime : $d^2 \geq n \geq d$.

Low rank estimation

Let

$$\mathcal{M}(k_0) = \{\theta : \operatorname{rank}(\theta) \leq k_0, \|\theta\|_\infty \leq 1\}.$$

There exists an adaptive estimator $\tilde{\theta}$ of $\theta \in \mathcal{M}(k_0)$ that achieves whp the minimax-optimal rate

$$\frac{\|\tilde{\theta} - \theta\|}{d} \lesssim \sqrt{\frac{k_0 d}{n}} := r(k_0).$$

where $\|.\|$ is the Frobenius norm, [Keshavan et al., 2009, Cai et al., 2010, Kolchinskii et al., 2011, Klopp and Gaiffas, 2015].

In terms of *estimation* of θ, these two models are equivalent.

Low rank estimation

Let

$$\mathcal{M}(k_0) = \{\theta : \text{rank}(\theta) \leq k_0, \|\theta\|_\infty \leq 1\}.$$

There exists an adaptive estimator $\tilde{\theta}$ of $\theta \in \mathcal{M}(k_0)$ that achieves whp the minimax-optimal rate

$$\frac{\|\tilde{\theta} - \theta\|}{d} \lesssim \sqrt{\frac{k_0 d}{n}} := r(k_0).$$

where $\|.\|$ is the Frobenius norm, [Keshavan et al., 2009, Cai et al., 2010, Kolchinskii et al., 2011, Klopp and Gaiffas, 2015].

Question : Can we construct a confidence set in $\|.\|$ norm with diameter scaling with $r(k_0)$ without the knowledge of k_0?

Adaptive and honest confidence sets

Adaptive and honest confidence set C_n

Let $\alpha > 0$. Set C_n that satisfies

- C_n covers θ, i.e.

$$\inf_{\theta \in \mathcal{M}(d)} P_\theta(\theta \in C_n) \geq 1 - \alpha. \qquad \text{[Honest Coverage]}$$

- For any $1 \leq k_0 \leq d$, the diameter of C_n satisfies for $\theta \in \mathcal{M}(k_0)$

$$|C_n| \lesssim_{whp} r(k_0). \qquad \text{[Optimal Diameter]}$$

Remark on adaptive and honest confidence sets

Consider, for $k_0 < d$, the testing problem :

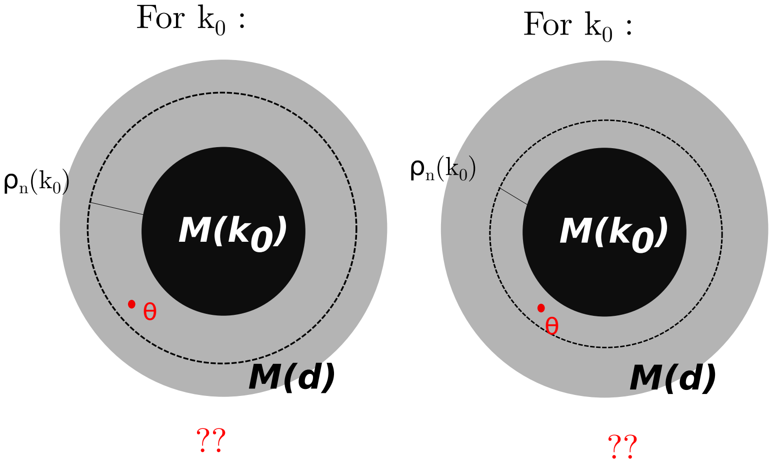
$$H_0 : \theta \in \mathcal{M}(k_0) \quad \text{vs} \quad H_1 : \theta \in \mathcal{M}(d), \|\theta - \mathcal{M}(k_0)\| \geq \rho.$$

Let $\rho_n(k_0)$ be the minimax-optimal testing rate.

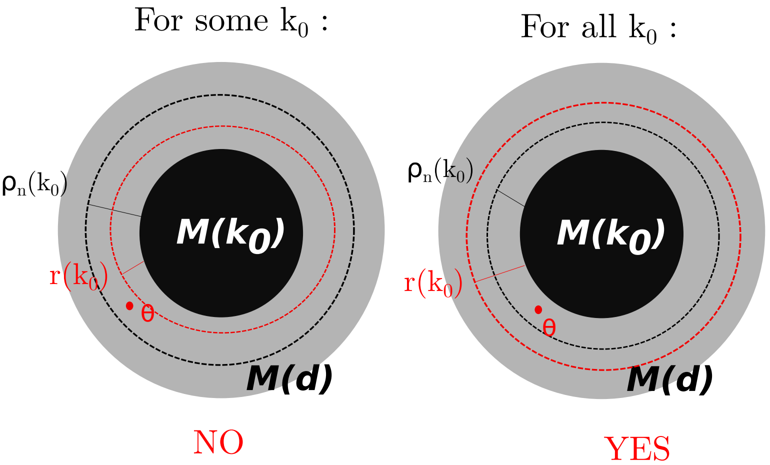
Honest and adaptive confidence sets exist over $\mathcal{M}(d)$
if and only if
for all $k_0 \leq d$, $\rho_n(k_0) \lesssim r(k_0)$.

Remark : This equivalence holds in many other settings [Low (1997), Cai and Low (2004), Robins and van der Vaart (2006), Hoffmann and Nickl (2011), Nickl and van de Geer(20013), C et al. (2015)].

Remark on adaptive and honest confidence sets



Remark on adaptive and honest confidence sets



Confidence sets : known $\mathbb{V}(\epsilon)$

Theorem [C, Klopp, Löffler, and Nickl (very soon)]

Let $d^2 \geq n$ and $\mathbb{V}(\epsilon)$ be fixed and known. Adaptive and honest confidence sets exist in both models (one needs in addition $n \geq d\log(d)$ in the Bernoulli model).

The two models are still equivalent regarding confidence sets with known $\mathbb{V}(\epsilon)$.

(Simplified) Idea of the proof

Let $\hat{\theta}$ be a minimax estimator s.t. $\mathrm{rank}(\hat{\theta}) \leq \mathrm{rank}(\theta)$ whp. Set

$T_n = \frac{\|Y - \mathcal{X}\hat{\theta}\|^2}{n} - \mathbb{V}(\epsilon)$, where $\mathcal{X}$ is the sampling operator.

We have whp and knowing $\tilde{\theta}$

$$|T_n - \|\theta - \hat{\theta}\|^2| \lesssim \frac{(\mathrm{rank}(\hat{\theta}) + 1)d}{n}.$$

If $\theta \in \mathcal{M}(k_0)$:

$$\hat{\theta} \in \mathcal{M}(k_0) \quad \text{and} \quad T_n \lesssim \frac{(k_0 + 1)d}{n} \simeq r_{k_0}^2.$$

If $r_{k_0} \lesssim \|\theta - \mathcal{M}(k_0)\|$:

$$\hat{\theta} \notin \mathcal{M}(k_0) \quad \text{or} \quad r_{k_0} - \frac{(k_0 + 1)d}{n} \lesssim r_{k_0}^2 \lesssim T_n.$$

This concludes the proof accepting the test if either $\hat{\theta} \in \mathcal{M}(k_0)$ or of $T_n \lesssim r_{k_0}^2$.

Confidence sets : unknown σ

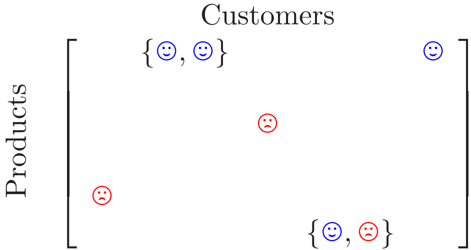
Lemma [C, Klopp, Löffler, and Nickl (very soon)]

Assume that $\mathbb{V}(\epsilon)$ is not known beforehand.

- **Trace Regression model** : If $d^2 \geq n \geq d$, adaptive and honest confidence sets exist.
- **Bernoulli Model** : Adaptive and honest confidence sets do not exist.

The two models are not equivalent in this case!

(Simplified) Idea of the proof : Trace Regression model



More than $\frac{n^2}{d^2} \wedge n$ entries sampled twices, i.e. N. Set $(\tilde{X}_i, Z_i, Z_i')_{i \le N}$

Replace T_n by

$$T_n = \frac{1}{N} \sum_{i \le N} (Z_i - \tilde{\mathcal{X}}\hat{\theta})(Z_i' - \tilde{\mathcal{X}}\hat{\theta}).$$

We have as before whp and knowing $\tilde{\theta}$

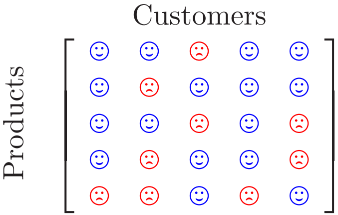
$$|T_n - \|\theta - \hat{\theta}\|^2| \lesssim \frac{(\mathrm{rank}(\hat{\theta}) + 1)d}{n}.$$

This concludes the proof.

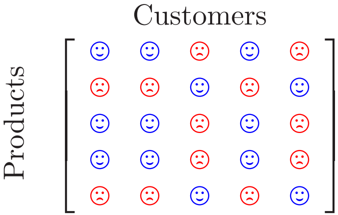
(Simplified) Idea of the proof : Bernoulli model

No entries sampled twice! First example : rank one

H_0 : Random opinions!



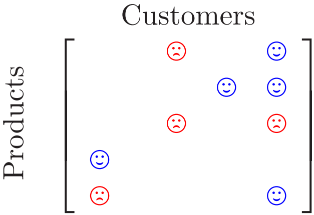
H_1 : Rank one opinions.



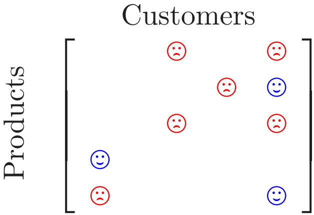
(Simplified) Idea of the proof : Bernoulli model

No entries sampled twice! First example : rank one

H_0 : Random opinions!



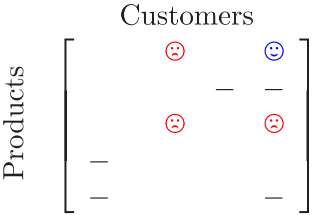
H_1 : Rank one opinions.



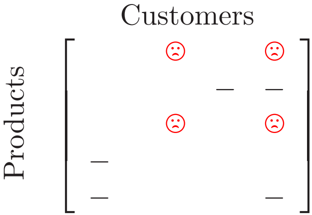
(Simplified) Idea of the proof : Bernoulli model

No entries sampled twice! First example : rank one

H_0 : Random opinions!



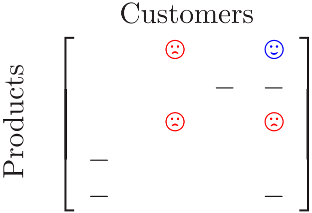
H_1 : Rank one opinions.



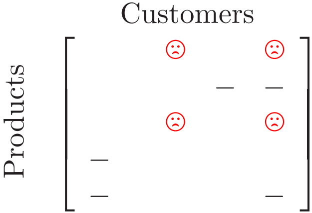
(Simplified) Idea of the proof : Bernoulli model

No entries sampled twice! First example : rank one

H_0 : Random opinions!



H_1 : Rank one opinions.

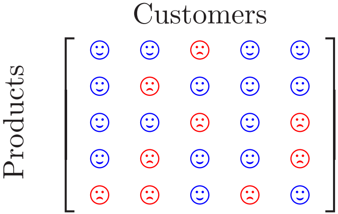


Less than $\frac{n^4}{d^4}$ such cycles whp $\rightarrow$ distinguishability only if $n \gg d$.

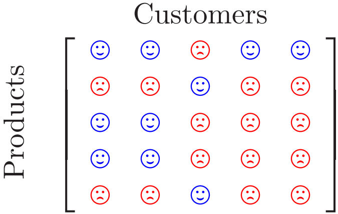
(Simplified) Idea of the proof : Bernoulli model

No entries sampled twice! General case : rank k

H_0 : Random opinions!



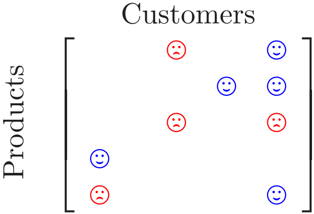
H_1 : Rank one opinions.



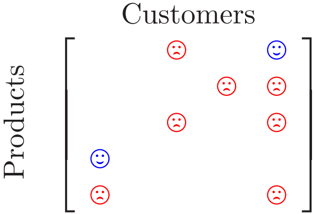
(Simplified) Idea of the proof : Bernoulli model

No entries sampled twice! General case : rank k

H_0 : Random opinions!



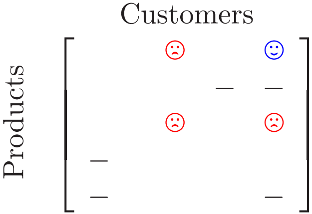
H_1 : Rank one opinions.



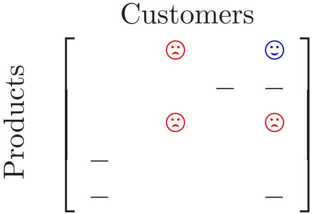
(Simplified) Idea of the proof : Bernoulli model

No entries sampled twice! General case : rank k

H_0 : Random opinions!



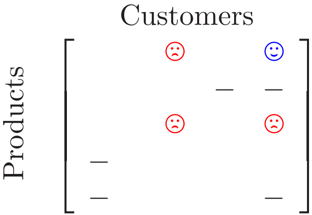
H_1 : Rank one opinions.



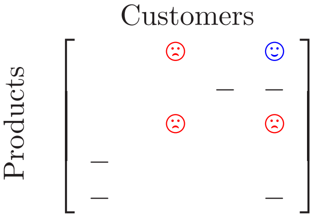
(Simplified) Idea of the proof : Bernoulli model

No entries sampled twice! General case : rank k

H_0 : Random opinions!



H_1 : Rank one opinions.



Less than $\frac{n^4}{d^4 k^3}$ *correct* cycles (taking rank groups into account)
$\rightarrow$ distinguishability only if $n \gg k^{3/4} d$.

Conclusion

- In general adaptive and honest confidence sets exist in the Trace Regression Model but not in the Bernoulli Model $\rightarrow$ these two very similar models are not equivalent for complete inference, although they are equivalent for estimation.
- Adaptive and honest confidence sets are linked to composite minimax testing rates. Results on whether they exist or not are therefore more subtle and model dependent than results for estimation.

THANK YOU!