Statistical Analysis of k-Nearest Neighbor Collaborative Recommendation

G. Biau, B. Cadre and L. Rouvière

University Paris VI — ENS Cachan — ENSAI

JOURNÉES STAR, OCTOBER 2009

Outline

Collaborative recommendation

- Collaborative recommendation is a Web information-filtering technique that typically
  - gathers information about your personal interests
  - compares your profile to other users with similar tastes
  - and then gives personalized recommendations.
- Examples include recommending books, people, restaurants, movies, CDs and news.
- Websites such as amazon.com, match.com, movielens.org and allmusic.com already have recommendation systems in operation.

Collaborative recommendation

- Collaborative recommendation is a Web information-filtering technique that typically
  - gathers information about your personal interests
  - compares your profile to other users with similar tastes
  - and then gives personalized recommendations.
- Examples include recommending books, people, restaurants, movies, CDs and news.
- Websites such as amazon.com, match.com, movielens.org and allmusic.com already have recommendation systems in operation.

Collaborative recommendation

- Collaborative recommendation is a Web information-filtering technique that typically
  - gathers information about your personal interests
  - compares your profile to other users with similar tastes
  - and then gives personalized recommendations.
- Examples include recommending books, people, restaurants, movies, CDs and news.
- Websites such as amazon.com, match.com, movielens.org and allmusic.com already have recommendation systems in operation.

Collaborative recommendation

- Collaborative recommendation is a Web information-filtering technique that typically
  - gathers information about your personal interests
  - compares your profile to other users with similar tastes
  - and then gives personalized recommendations.
- Examples include recommending books, people, restaurants, movies, CDs and news.
- Websites such as amazon.com, match.com, movielens.org and allmusic.com already have recommendation systems in operation.

Collaborative recommendation

- Collaborative recommendation is a Web information-filtering technique that typically
  - gathers information about your personal interests
  - compares your profile to other users with similar tastes
  - and then gives personalized recommendations.
- Examples include recommending books, people, restaurants, movies, CDs and news.
- Websites such as amazon.com, match.com, movielens.org and allmusic.com already have recommendation systems in operation.

Collaborative recommendation

- Collaborative recommendation is a Web information-filtering technique that typically
  - ▷ gathers information about your personal interests
  - ▷ compares your profile to other users with similar tastes
  - ▷ and then gives personalized recommendations.
- Examples include recommending books, people, restaurants, movies, CDs and news.
- Websites such as amazon.com, match.com, movielens.org and allmusic.com already have recommendation systems in operation.

Users and items

- Collaborative systems deal with two types of variables: users and items.
- The problem: Estimate ratings for items that have not yet been consumed by a user.
- The recommendation process typically starts by asking users a series of questions.
- Personal ratings are then collected in a matrix.

Based on this prior information, the recommendation engine must be able to automatically furnish ratings of as-yet unrated items.

Users and items

- Collaborative systems deal with two types of variables: users and items.
- The problem: Estimate ratings for items that have not yet been consumed by a user.
- The recommendation process typically starts by asking users a series of questions.
- Personal ratings are then collected in a matrix.

Based on this prior information, the recommendation engine must be able to automatically furnish ratings of as-yet unrated items.

Users and items

- Collaborative systems deal with two types of variables: users and items.
- The problem: Estimate ratings for items that have not yet been consumed by a user.
- The recommendation process typically starts by asking users a series of questions.
- Personal ratings are then collected in a matrix.

Based on this prior information, the recommendation engine must be able to automatically furnish ratings of as-yet unrated items.

Users and items

- Collaborative systems deal with two types of variables: users and items.
- The problem: Estimate ratings for items that have not yet been consumed by a user.
- The recommendation process typically starts by asking users a series of questions.
- Personal ratings are then collected in a matrix.

Based on this prior information, the recommendation engine must be able to automatically furnish ratings of as-yet unrated items.

Users and items

- Collaborative systems deal with two types of variables: users and items.
- The problem: Estimate ratings for items that have not yet been consumed by a user.
- The recommendation process typically starts by asking users a series of questions.
- Personal ratings are then collected in a matrix.

Based on this prior information, the recommendation engine must be able to automatically furnish ratings of as-yet unrated items.

Users and items

- Collaborative systems deal with two types of variables: users and items.
- The problem: Estimate ratings for items that have not yet been consumed by a user.
- The recommendation process typically starts by asking users a series of questions.
- Personal ratings are then collected in a matrix.

Based on this prior information, the recommendation engine must be able to automatically furnish ratings of as-yet unrated items.

Users and items

- Collaborative systems deal with two types of variables: users and items.
- The problem: Estimate ratings for items that have not yet been consumed by a user.
- The recommendation process typically starts by asking users a series of questions.
- Personal ratings are then collected in a matrix.

Based on this prior information, the recommendation engine must be able to automatically furnish ratings of as-yet unrated items.

Example

	Rocky	Platoon	Rambo	Rio Bravo	Star wars	Titanic
Jim	NA	6	7	8	9	NA
James	3	NA	10	NA	5	7
Steve	7	NA	1	NA	6	NA
Mary	NA	7	1	NA	5	6
John	NA	7	NA	NA	3	1
Lucy	3	10	2	7	NA	4
Stan	NA	7	NA	NA	1	NA
Johanna	4	5	NA	8	3	9

Example

	Rocky	Platoon	Rambo	Rio Bravo	Star wars	Titanic
Jim	NA	6	7	8	9	NA
James	3	NA	10	NA	5	7
Steve	7	NA	1	NA	6	NA
Mary	NA	7	1	NA	5	6
John	NA	7	NA	NA	3	1
Lucy	3	10	2	7	NA	4
Stan	NA	7	NA	NA	1	NA
Johanna	4	5	NA	8	3	9
Bob	NA	3	3	4	5	?

State of the art

- A number of practical methods have been proposed, including
  - ▷ machine learning-oriented techniques (e.g., Abernethy et al., 2009)
  - ▷ statistical approaches (e.g., Sarwar et al., 2001)
  - ▷ and numerous other ad hoc rules (Adomavicius and Tuzhilin, 2005).
- The similarity measure assessing proximity between users is typically based on a correlation or cosine-type approach.

State of the art

- A number of practical methods have been proposed, including
  - machine learning-oriented techniques (e.g., Abernethy et al., 2009)
  - statistical approaches (e.g., Sarwar et al., 2001)
  - and numerous other ad hoc rules (Adomavicius and Tuzhilin, 2005).
- The similarity measure assessing proximity between users is typically based on a correlation or cosine-type approach.

State of the art

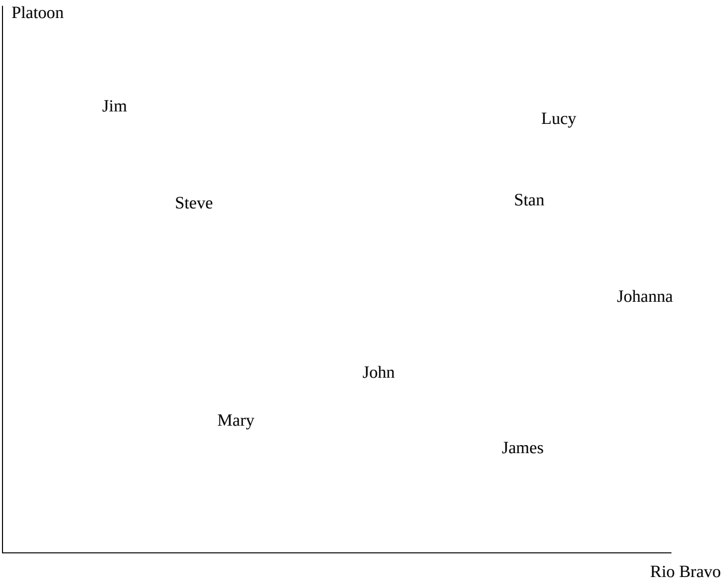
- A number of practical methods have been proposed, including
  - ▷ machine learning-oriented techniques (e.g., Abernethy et al., 2009)
  - ▷ statistical approaches (e.g., Sarwar et al., 2001)
  - ▷ and numerous other ad hoc rules (Adomavicius and Tuzhilin, 2005).
- The similarity measure assessing proximity between users is typically based on a correlation or cosine-type approach.

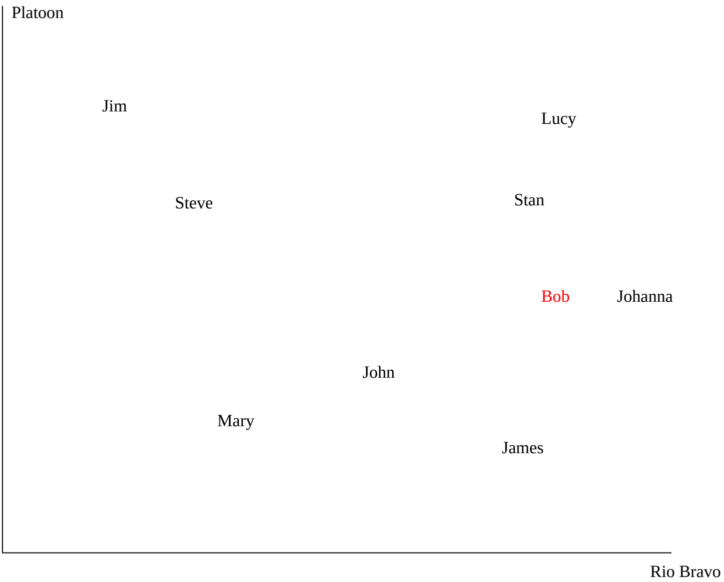
State of the art

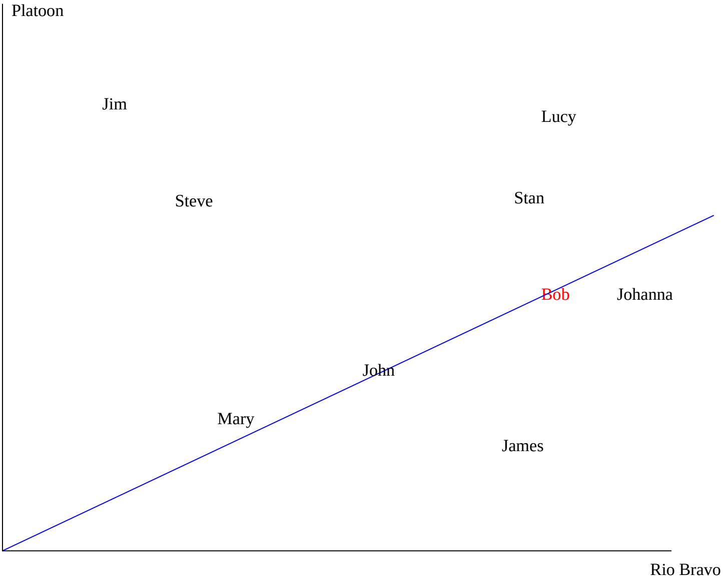
- A number of practical methods have been proposed, including
  - machine learning-oriented techniques (e.g., Abernethy et al., 2009)
  - statistical approaches (e.g., Sarwar et al., 2001)
  - and numerous other ad hoc rules (Adomavicius and Tuzhilin, 2005).
- The similarity measure assessing proximity between users is typically based on a correlation or cosine-type approach.

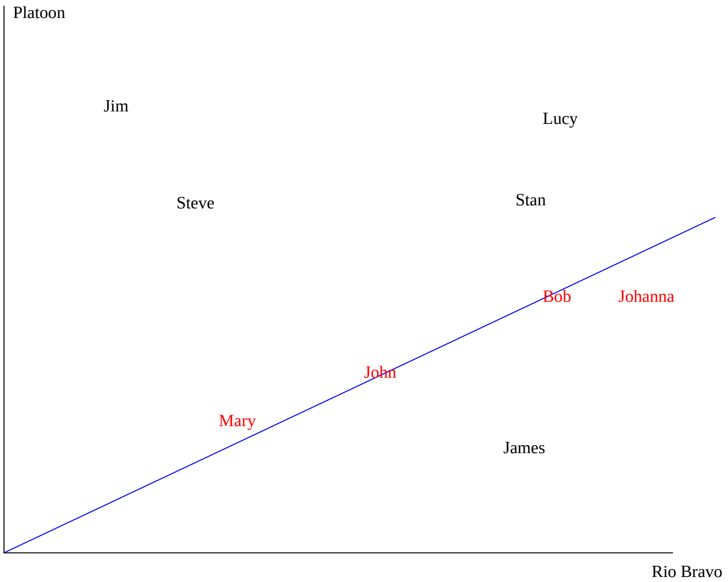
State of the art

- A number of practical methods have been proposed, including
  - ▷ machine learning-oriented techniques (e.g., `Abernethy et al., 2009`)
  - ▷ statistical approaches (e.g., `Sarwar et al., 2001`)
  - ▷ and numerous other ad hoc rules (`Adomavicius and Tuzhilin, 2005`).
- The similarity measure assessing proximity between users is typically based on a correlation or cosine-type approach.


Platoon
Jim
Lucy
Steve
Stan
Johanna
John
Mary
James
Rio Bravo


Platoon
Jim
Lucy
Steve
Stan
Bob
Johanna
John
Mary
James
Rio Bravo


Platoon
Jim
Lucy
Steve
Stan
Bob
Johanna
John
Mary
James
Rio Bravo


Platoon
Jim
Lucy
Steve
Stan
Bob
Johanna
John
Mary
James
Rio Bravo

Our mission

- Despite wide-ranging literature, very little is known about the statistical properties of recommendation systems.
- In fact, no clear probabilistic model even exists.
- To provide an initial contribution to this, we propose
  - to set out a general stochastic model for collaborative recommendation
  - and analyze its asymptotic performance as the number of users grows.

Our mission

- Despite wide-ranging literature, very little is known about the statistical properties of recommendation systems.
- In fact, no clear probabilistic model even exists.
- To provide an initial contribution to this, we propose
  - to set out a general stochastic model for collaborative recommendation
  - and analyze its asymptotic performance as the number of users grows.

Our mission

- Despite wide-ranging literature, very little is known about the statistical properties of recommendation systems.
- In fact, no clear probabilistic model even exists.
- To provide an initial contribution to this, we propose
  - to set out a general stochastic model for collaborative recommendation
  - and analyze its asymptotic performance as the number of users grows.

Our mission

- Despite wide-ranging literature, very little is known about the statistical properties of recommendation systems.
- In fact, no clear probabilistic model even exists.
- To provide an initial contribution to this, we propose
  - to set out a general stochastic model for collaborative recommendation
  - and analyze its asymptotic performance as the number of users grows.

Our mission

- Despite wide-ranging literature, very little is known about the statistical properties of recommendation systems.
- In fact, no clear probabilistic model even exists.
- To provide an initial contribution to this, we propose
  - to set out a general stochastic model for collaborative recommendation
  - and analyze its asymptotic performance as the number of users grows.

Ratings matrix and new users

- Suppose that there are $d+1$ possible items, n users in the ratings matrix.
- Users' ratings take values in the set $(\{0\} \cup [1, s])^{d+1}$.
- A new user Bob reveals some of his preferences for the first time, rating some of the first d items but not the $(d+1)$th.

We want to design a strategy to predict Bob's rating of Titanic, using

1. Bob's ratings of some of the other d movies and
2. the ratings matrix.

Ratings matrix and new users

- Suppose that there are $d+1$ possible items, n users in the ratings matrix.
- Users' ratings take values in the set $(\{0\} \cup [1, s])^{d+1}$.
- A new user Bob reveals some of his preferences for the first time, rating some of the first d items but not the $(d+1)$th.

We want to design a strategy to predict Bob's rating of Titanic, using

1. Bob's ratings of some of the other d movies and
2. the ratings matrix.

Ratings matrix and new users

- Suppose that there are $d+1$ possible items, n users in the ratings matrix.
- Users' ratings take values in the set $(\{0\} \cup [1, s])^{d+1}$.
- A new user Bob reveals some of his preferences for the first time, rating some of the first d items but not the $(d+1)$th.

We want to design a strategy to predict Bob's rating of Titanic, using

1. Bob's ratings of some of the other d movies and
2. the ratings matrix.

Ratings matrix and new users

- Suppose that there are $d+1$ possible items, n users in the ratings matrix.
- Users' ratings take values in the set $(\{0\} \cup [1, s])^{d+1}$.
- A new user Bob reveals some of his preferences for the first time, rating some of the first d items but not the $(d+1)$th.

We want to design a strategy to predict Bob's rating of Titanic, using

1. Bob's ratings of some of the other d movies and
2. the ratings matrix.

Ratings matrix and new users

- Suppose that there are $d+1$ possible items, n users in the ratings matrix.
- Users' ratings take values in the set $(\{0\} \cup [1, s])^{d+1}$.
- A new user Bob reveals some of his preferences for the first time, rating some of the first d items but not the $(d+1)$th.

We want to design a strategy to predict Bob's rating of Titanic, using

1. Bob's ratings of some of the other d movies and
2. the ratings matrix.

Ratings matrix and new users

- Suppose that there are $d+1$ possible items, n users in the ratings matrix.
- Users' ratings take values in the set $(\{0\} \cup [1, s])^{d+1}$.
- A new user Bob reveals some of his preferences for the first time, rating some of the first d items but not the $(d+1)$th.

We want to design a strategy to predict Bob's rating of Titanic, using

1. Bob's ratings of some of the other d movies and
2. the ratings matrix.

A model for the new user

- We model the preferences of Bob by a random vector $(\mathbf{X}, Y)$ taking values in the set $[1, s]^d \times [1, s]$.
- In fact, we do not observe the variable $\mathbf{X}$, but instead some "masked" version of it $\mathbf{X}^\star$.
- The random variable $\mathbf{X}^\star = (X_1^\star, \ldots, X_d^\star)$ is naturally defined by

$$X_j^\star = \begin{cases} X_j & \text{if } j \in M \\ 0 & \text{otherwise,} \end{cases}$$

where M stands for some non-empty random subset of $\{1, \ldots, d\}$.

Bob	NA	3	3	4	5	?

$M = \{2, 3, 4, 5\}$ $\quad$ $\mathbf{X}^\star = (0, 3, 3, 4, 5)$.

A model for the new user

- We model the preferences of Bob by a random vector $(\mathbf{X}, Y)$ taking values in the set $[1,s]^d \times [1,s]$.
- In fact, we do not observe the variable $\mathbf{X}$, but instead some "masked" version of it $\mathbf{X}^\star$.
- The random variable $\mathbf{X}^\star = (X_1^\star, \ldots, X_d^\star)$ is naturally defined by

$$X_j^\star = \begin{cases} X_j & \text{if } j \in M \\ 0 & \text{otherwise,} \end{cases}$$

where M stands for some non-empty random subset of $\{1, \ldots, d\}$.

Bob	NA	3	3	4	5	?

$M = \{2,3,4,5\}$ $\qquad$ $\mathbf{X}^\star = (0,3,3,4,5).$

A model for the new user

- We model the preferences of Bob by a random vector $(\mathbf{X}, Y)$ taking values in the set $[1, s]^d \times [1, s]$.
- In fact, we do not observe the variable $\mathbf{X}$, but instead some "masked" version of it $\mathbf{X}^\star$.
- The random variable $\mathbf{X}^\star = (X_1^\star, \ldots, X_d^\star)$ is naturally defined by

$$X_j^\star = \begin{cases} X_j & \text{if } j \in M \\ 0 & \text{otherwise,} \end{cases}$$

where M stands for some non-empty random subset of $\{1, \ldots, d\}$.

Bob	NA	3	3	4	5	?

$$M = \{2, 3, 4, 5\} \qquad \mathbf{X}^\star = (0, 3, 3, 4, 5).$$

A first model for the ratings matrix

- The preferences of users already in the database will be represented by independent random pairs $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ from the distribution $(\mathbf{X}, Y)$.
- A first idea for dealing with potential non-responses is to consider in place of $\mathbf{X}_i$ its masked version $\widetilde{\mathbf{X}}_i = (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{id})$ defined by

$$\widetilde{X}_{ij} = \begin{cases} X_{ij} & \text{if } j \in M_i \cap M \\ 0 & \text{otherwise.} \end{cases}$$

Drawback

As time goes by, each user in the database may reveal online more and more preferences.

A first model for the ratings matrix

- The preferences of users already in the database will be represented by independent random pairs $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ from the distribution $(\mathbf{X}, Y)$.

- A first idea for dealing with potential non-responses is to consider in place of $\mathbf{X}_i$ its masked version $\widetilde{\mathbf{X}}_i = (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{id})$ defined by

$$\widetilde{X}_{ij} = \begin{cases} X_{ij} & \text{if } j \in M_i \cap M \\ 0 & \text{otherwise.} \end{cases}$$

Drawback

As time goes by, each user in the database may reveal online more and more preferences.

A first model for the ratings matrix

- The preferences of users already in the database will be represented by independent random pairs $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ from the distribution $(\mathbf{X}, Y)$.
- A first idea for dealing with potential non-responses is to consider in place of $\mathbf{X}_i$ its masked version $\widetilde{\mathbf{X}}_i = (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{id})$ defined by

$$\widetilde{X}_{ij} = \begin{cases} X_{ij} & \text{if } j \in M_i \cap M \\ 0 & \text{otherwise.} \end{cases}$$

Drawback

As time goes by, each user in the database may reveal online more and more preferences.

A more pertinent model

- At time 1, there is only one user in the database, and the subset of items he decides to rate is modeled by a random variable M_1^1.
- At time 2, a new user enters the game and reveals his preferences according to a variable M_2^1, with the same distribution as M_1^1.
- At the same time, user 1 may update his list of preferences, modeled by a random variable M_1^2 satisfying $M_1^1 \subset M_1^2$.
- And so on...

A more pertinent model

- At time 1, there is only one user in the database, and the subset of items he decides to rate is modeled by a random variable M_1^1.
- At time 2, a new user enters the game and reveals his preferences according to a variable M_2^1, with the same distribution as M_1^1.
- At the same time, user 1 may update his list of preferences, modeled by a random variable M_1^2 satisfying $M_1^1 \subset M_1^2$.
- And so on...

A more pertinent model

- At time 1, there is only one user in the database, and the subset of items he decides to rate is modeled by a random variable M_1^1.
- At time 2, a new user enters the game and reveals his preferences according to a variable M_2^1, with the same distribution as M_1^1.
- At the same time, user 1 may update his list of preferences, modeled by a random variable M_1^2 satisfying $M_1^1 \subset M_1^2$.
- And so on...

A more pertinent model

- At time 1, there is only one user in the database, and the subset of items he decides to rate is modeled by a random variable M_1^1.
- At time 2, a new user enters the game and reveals his preferences according to a variable M_2^1, with the same distribution as M_1^1.
- At the same time, user 1 may update his list of preferences, modeled by a random variable M_1^2 satisfying $M_1^1 \subset M_1^2$.
- And so on...

Preference updating

	Time 1	Time 2	...	Time i	...	Time n
User 1	M_1^1	M_1^2	...	M_1^i	...	M_1^n
User 2		M_2^1	...	M_2^{i-1}	...	M_2^{n-1}
$\vdots$			$\ddots$	$\vdots$	$\vdots$	$\vdots$
User i				M_i^1	...	M_i^{n+1-i}
$\vdots$					$\ddots$	$\vdots$
User n						M_n^1

Assumptions

1. The distribution of $(M_i^n)_{n\geq 1}$ is independent of i. It is therefore distributed as a generic random sequence $(M^n)_{n\geq 1}$, satisfying $M^1 \neq \emptyset$ and $M^n \subset M^{n+1}$ for all $n \geq 1$.

2. There exists a random integer n_0 such that $M^{n_0} = \{1, \ldots, d\}$.

Preference updating

	Time 1	Time 2	...	Time i	...	Time n
User 1	M_1^1	M_1^2	...	M_1^i	...	M_1^n
User 2		M_2^1	...	M_2^{i-1}	...	M_2^{n-1}
$\vdots$			$\ddots$	$\vdots$	$\vdots$	$\vdots$
User i				M_i^1	...	M_i^{n+1-i}
$\vdots$					$\ddots$	$\vdots$
User n						M_n^1

Assumptions

1. The distribution of $(M_i^n)_{n\geq 1}$ is independent of i. It is therefore distributed as a generic random sequence $(M^n)_{n\geq 1}$, satisfying $M^1 \neq \emptyset$ and $M^n \subset M^{n+1}$ for all $n \geq 1$.
2. There exists a random integer n_0 such that $M^{n_0} = \{1, \ldots, d\}$.

Preference updating

	Time 1	Time 2	...	Time i	...	Time n
User 1	M_1^1	M_1^2	...	M_1^i	...	M_1^n
User 2		M_2^1	...	M_2^{i-1}	...	M_2^{n-1}
$\vdots$			$\ddots$	$\vdots$	$\vdots$	$\vdots$
User i				M_i^1	...	M_i^{n+1-i}
$\vdots$					$\ddots$	$\vdots$
User n						M_n^1

Assumptions

1. The distribution of $(M_i^n)_{n\geq 1}$ is independent of i. It is therefore distributed as a generic random sequence $(M^n)_{n\geq 1}$, satisfying $M^1 \neq \emptyset$ and $M^n \subset M^{n+1}$ for all $n \geq 1$.
2. There exists a random integer n_0 such that $M^{n_0} = \{1, \ldots, d\}$.

The sequential model

- We let the masked version $\mathbf{X}_i^{(n)}$ of $\mathbf{X}_i$ be defined as

$$X_{ij}^{(n)} = \begin{cases} X_{ij} & \text{if } j \in M_i^{n+1-i} \cap M \\ 0 & \text{otherwise.} \end{cases}$$

- We denote by $(\mathcal{R}_n)_{n\geq 1}$ the subset of users who have already provided information about Titanic at time n. It satisfies $\mathcal{R}_n \subset \mathcal{R}_{n+1}$.

The statistical problem

Estimate the regression function $\eta(\mathbf{x}^\star) = \mathbb{E}[Y|\mathbf{X}^\star = \mathbf{x}^\star]$, based on the database observations $(\mathbf{X}_1^{(n)}, Y_1), \ldots, (\mathbf{X}_n^{(n)}, Y_n)$.

The sequential model

- We let the masked version $\mathbf{X}_i^{(n)}$ of $\mathbf{X}_i$ be defined as

$$X_{ij}^{(n)} = \begin{cases} X_{ij} & \text{if } j \in M_i^{n+1-i} \cap M \\ 0 & \text{otherwise.} \end{cases}$$

- We denote by $(\mathcal{R}_n)_{n\geq 1}$ the subset of users who have already provided information about Titanic at time n. It satisfies $\mathcal{R}_n \subset \mathcal{R}_{n+1}$.

The statistical problem

Estimate the regression function $\eta(\mathbf{x}^\star) = \mathbb{E}[Y|\mathbf{X}^\star = \mathbf{x}^\star]$, based on the database observations $(\mathbf{X}_1^{(n)}, Y_1), \ldots, (\mathbf{X}_n^{(n)}, Y_n)$.

The sequential model

- We let the masked version $\mathbf{X}_i^{(n)}$ of $\mathbf{X}_i$ be defined as

$$X_{ij}^{(n)} = \begin{cases} X_{ij} & \text{if } j \in M_i^{n+1-i} \cap M \\ 0 & \text{otherwise.} \end{cases}$$

- We denote by $(\mathcal{R}_n)_{n\geq 1}$ the subset of users who have already provided information about Titanic at time n. It satisfies $\mathcal{R}_n \subset \mathcal{R}_{n+1}$.

The statistical problem

Estimate the regression function $\eta(\mathbf{x}^\star) = \mathbb{E}[Y|\mathbf{X}^\star = \mathbf{x}^\star]$, based on the database observations $(\mathbf{X}_1^{(n)}, Y_1), \ldots, (\mathbf{X}_n^{(n)}, Y_n)$.

The cosine-type k_n-NN regression method

- The method estimate $\eta(\mathbf{x}^\star)$ by a local averaging over those Y_i for which
  1. $\mathbf{X}_i^{(n)}$ is "close" to $\mathbf{x}^\star$ and
  2. $i \in \mathcal{R}_n$, that is, we effectively "see" the rating Y_i.
- The closeness between users is assessed by a cosine-type similarity, defined by

$$\bar{S}(\mathbf{x}, \mathbf{x}') = \frac{\sum_{j \in \mathcal{J}} x_j x'_j}{\sqrt{\sum_{j \in \mathcal{J}} x_j^2} \sqrt{\sum_{j \in \mathcal{J}} x_j'^2}},$$

  where $\mathcal{J} = \{j \in \{1, \ldots, d\} : x_j \neq 0 \text{ and } x'_j \neq 0\}$.
- If $\mathcal{J} = \{1, \ldots, d\}$ then $\bar{S}(\mathbf{x}, \mathbf{x}') = \cos(\mathbf{x}, \mathbf{x}')$.

The cosine-type k_n-NN regression method

- The method estimate $\eta(\mathbf{x}^\star)$ by a local averaging over those Y_i for which
  1. $\mathbf{X}_i^{(n)}$ is "close" to $\mathbf{x}^\star$ and
  2. $i \in \mathcal{R}_n$, that is, we effectively "see" the rating Y_i.
- The closeness between users is assessed by a cosine-type similarity, defined by

$$\bar{S}(\mathbf{x}, \mathbf{x}') = \frac{\sum_{j \in \mathcal{J}} x_j x_j'}{\sqrt{\sum_{j \in \mathcal{J}} x_j^2} \sqrt{\sum_{j \in \mathcal{J}} x_j'^2}},$$

  where $\mathcal{J} = \{j \in \{1, \ldots, d\} : x_j \neq 0 \text{ and } x_j' \neq 0\}$.
- If $\mathcal{J} = \{1, \ldots, d\}$ then $\bar{S}(\mathbf{x}, \mathbf{x}') = \cos(\mathbf{x}, \mathbf{x}')$.

The cosine-type k_n-NN regression method

- The method estimate $\eta(\mathbf{x}^\star)$ by a local averaging over those Y_i for which
  1. $\mathbf{X}_i^{(n)}$ is "close" to $\mathbf{x}^\star$ and
  2. $i \in \mathcal{R}_n$, that is, we effectively "see" the rating Y_i.
- The closeness between users is assessed by a cosine-type similarity, defined by

$$\bar{S}(\mathbf{x}, \mathbf{x}') = \frac{\sum_{j\in\mathcal{J}} x_j x'_j}{\sqrt{\sum_{j\in\mathcal{J}} x_j^2}\sqrt{\sum_{j\in\mathcal{J}} x_j'^2}},$$

  where $\mathcal{J} = \{j \in \{1, \ldots, d\} : x_j \neq 0 \text{ and } x'_j \neq 0\}$.
- If $\mathcal{J} = \{1, \ldots, d\}$ then $\bar{S}(\mathbf{x}, \mathbf{x}') = \cos(\mathbf{x}, \mathbf{x}')$.

The cosine-type k_n-NN regression method

- The method estimate $\eta(\mathbf{x}^\star)$ by a local averaging over those Y_i for which
  1. $\mathbf{X}_i^{(n)}$ is "close" to $\mathbf{x}^\star$ and
  2. $i \in \mathcal{R}_n$, that is, we effectively "see" the rating Y_i.
- The closeness between users is assessed by a cosine-type similarity, defined by

$$\bar{S}(\mathbf{x}, \mathbf{x}') = \frac{\sum_{j \in \mathcal{J}} x_j x_j'}{\sqrt{\sum_{j \in \mathcal{J}} x_j^2} \sqrt{\sum_{j \in \mathcal{J}} x_j'^2}},$$

  where $\mathcal{J} = \{j \in \{1, \ldots, d\} : x_j \neq 0 \text{ and } x_j' \neq 0\}$.
- If $\mathcal{J} = \{1, \ldots, d\}$ then $\bar{S}(\mathbf{x}, \mathbf{x}') = \cos(\mathbf{x}, \mathbf{x}')$.

The cosine-type k_n-NN regression method

- The method estimate $\eta(\mathbf{x}^\star)$ by a local averaging over those Y_i for which
  1. $\mathbf{X}_i^{(n)}$ is "close" to $\mathbf{x}^\star$ and
  2. $i \in \mathcal{R}_n$, that is, we effectively "see" the rating Y_i.
- The closeness between users is assessed by a cosine-type similarity, defined by

$$\bar{S}(\mathbf{x}, \mathbf{x}') = \frac{\sum_{j \in \mathcal{J}} x_j x_j'}{\sqrt{\sum_{j \in \mathcal{J}} x_j^2}\sqrt{\sum_{j \in \mathcal{J}} x_j'^2}},$$

where $\mathcal{J} = \{j \in \{1, \ldots, d\} : x_j \neq 0 \text{ and } x_j' \neq 0\}$.

- If $\mathcal{J} = \{1, \ldots, d\}$ then $\bar{S}(\mathbf{x}, \mathbf{x}') = \cos(\mathbf{x}, \mathbf{x}')$.

Example

$$\bar{S}(\text{Bob}, \text{Jim}) = \bar{S}((0, 3, 3, 4, 5), (0, 6, 7, 8, 9)) \approx 0.99,$$

whereas

$$\bar{S}(\text{Bob}, \text{Lucy}) = \bar{S}((0, 3, 3, 4, 5), (3, 10, 2, 7, 0)) \approx 0.89.$$

A key observation

If $M \subset M_i^{n+1-i}$, the positive real number y which maximizes the similarity between $(\mathbf{x}^\star, y)$ and $(\mathbf{X}_i^\star, Y_i)$ is given by

$$y = \frac{\|\mathbf{x}^\star\|}{\|\mathbf{X}_i^\star\| \cos(\mathbf{x}^\star, \mathbf{X}_i^\star)} Y_i.$$

Example

$$\bar{S}(\text{Bob}, \text{Jim}) = \bar{S}((0, 3, 3, 4, 5), (0, 6, 7, 8, 9)) \approx 0.99,$$

whereas

$$\bar{S}(\text{Bob}, \text{Lucy}) = \bar{S}((0, 3, 3, 4, 5), (3, 10, 2, 7, 0)) \approx 0.89.$$

A key observation

If $M \subset M_i^{n+1-i}$, the positive real number y which maximizes the similarity between $(\mathbf{x}^\star, y)$ and $(\mathbf{X}_i^\star, Y_i)$ is given by

$$y = \frac{\|\mathbf{x}^\star\|}{\|\mathbf{X}_i^\star\| \cos(\mathbf{x}^\star, \mathbf{X}_i^\star)} Y_i.$$

The estimate

- This suggests the regression estimate

$$\eta_n(\mathbf{x}^\star) = \|\mathbf{x}^\star\| \sum_{i \in \mathcal{R}_n} W_{ni}(\mathbf{x}^\star) \frac{Y_i}{\|\mathbf{X}_i^{(n)}\|},$$

where

$$W_{ni}(\mathbf{x}^\star) = \begin{cases} 1/k_n & \text{if } \mathbf{X}_i^{(n)} \text{ is among the } k_n\text{-MS of } \mathbf{x}^\star \text{ in } \{\mathbf{X}_i^{(n)}, i \in \mathcal{R}_n\} \\ 0 & \text{otherwise.} \end{cases}$$

- The weights are computed according to the penalized similarity

$$S\left(\mathbf{x}^\star, \mathbf{X}_i^{(n)}\right) = p_i^{(n)} \bar{S}\left(\mathbf{x}^\star, \mathbf{X}_i^{(n)}\right), \quad \text{with } p_i^{(n)} = \frac{|M_i^{n+1-i} \cap M|}{|M|}.$$

The estimate

- This suggests the regression estimate

$$\eta_n(\mathbf{x}^\star) = \|\mathbf{x}^\star\| \sum_{i \in \mathcal{R}_n} W_{ni}(\mathbf{x}^\star) \frac{Y_i}{\|\mathbf{X}_i^{(n)}\|},$$

where

$$W_{ni}(\mathbf{x}^\star) = \begin{cases} 1/k_n & \text{if } \mathbf{X}_i^{(n)} \text{ is among the } k_n\text{-MS of } \mathbf{x}^\star \text{ in } \{\mathbf{X}_i^{(n)}, i \in \mathcal{R}_n\} \\ 0 & \text{otherwise.} \end{cases}$$

- The weights are computed according to the penalized similarity

$$S\left(\mathbf{x}^\star, \mathbf{X}_i^{(n)}\right) = p_i^{(n)} \bar{S}\left(\mathbf{x}^\star, \mathbf{X}_i^{(n)}\right), \quad \text{with } p_i^{(n)} = \frac{|M_i^{n+1-i} \cap M|}{|M|}.$$

The regression function

Theorem

Suppose that $\eta_n(\mathbf{X}^\star) \to \eta(\mathbf{X}^\star)$ *in probability as* $n \to \infty$. *Then*

$$\eta(\mathbf{X}^\star) = \|\mathbf{X}^\star\| \, \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|} \Big| \frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} \right] \quad a.s.$$

Fundamental requirement

The regression function $\eta(\mathbf{x}^\star)$ has the special form

$$\eta(\mathbf{x}^\star) = \|\mathbf{x}^\star\| \varphi(\mathbf{x}^\star), \quad \text{where} \quad \varphi(\mathbf{x}^\star) = \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|} \Big| \frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} = \frac{\mathbf{x}^\star}{\|\mathbf{x}^\star\|} \right].$$

In particular, if $\tilde{\mathbf{x}}^\star = \lambda \mathbf{x}^\star$ with $\lambda > 0$, then $\eta(\tilde{\mathbf{x}}^\star) = \lambda \eta(\mathbf{x}^\star)$.

The regression function

Theorem

Suppose that $\eta_n(\mathbf{X}^\star) \to \eta(\mathbf{X}^\star)$ *in probability as* $n \to \infty$. *Then*

$$\eta(\mathbf{X}^\star) = \|\mathbf{X}^\star\| \, \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|} \Big| \frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} \right] \quad a.s.$$

Fundamental requirement

The regression function $\eta(\mathbf{x}^\star)$ has the special form

$$\eta(\mathbf{x}^\star) = \|\mathbf{x}^\star\| \varphi(\mathbf{x}^\star), \quad \text{where} \quad \varphi(\mathbf{x}^\star) = \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|} \Big| \frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} = \frac{\mathbf{x}^\star}{\|\mathbf{x}^\star\|} \right].$$

In particular, if $\tilde{\mathbf{x}}^\star = \lambda \mathbf{x}^\star$ with $\lambda > 0$, then $\eta(\tilde{\mathbf{x}}^\star) = \lambda \eta(\mathbf{x}^\star)$.

Consistency

Theorem

Suppose that $k_n \to \infty$, $|\mathcal{R}_n| \to \infty$ *a.s. and* $\mathbb{E}[k_n/|\mathcal{R}_n|] \to 0$ *as* $n \to \infty$. *Then*

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| \to 0 \quad \text{as } n \to \infty.$$

Thus, to achieve consistency, the number of nearest neighbors k_n should

- tend to infinity
- but be small with respect to the users who have already rated the item of interest.

Consistency

Theorem

Suppose that $k_n \to \infty$, $|\mathcal{R}_n| \to \infty$ *a.s. and* $\mathbb{E}[k_n/|\mathcal{R}_n|] \to 0$ *as* $n \to \infty$. *Then*

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| \to 0 \quad \text{as } n \to \infty.$$

Thus, to achieve consistency, the number of nearest neighbors k_n should

- tend to infinity
- but be small with respect to the users who have already rated the item of interest.

Consistency

Theorem

Suppose that $k_n \to \infty$, $|\mathcal{R}_n| \to \infty$ a.s. and $\mathbb{E}[k_n/|\mathcal{R}_n|] \to 0$ as $n \to \infty$. Then

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| \to 0 \quad \text{as } n \to \infty.$$

Thus, to achieve consistency, the number of nearest neighbors k_n should

- tend to infinity
- but be small with respect to the users who have already rated the item of interest.

Consistency

Theorem

Suppose that $k_n \to \infty$, $|\mathcal{R}_n| \to \infty$ *a.s. and* $\mathbb{E}[k_n/|\mathcal{R}_n|] \to 0$ *as* $n \to \infty$. *Then*

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| \to 0 \quad \text{as } n \to \infty.$$

Thus, to achieve consistency, the number of nearest neighbors k_n should

- tend to infinity
- but be small with respect to the users who have already rated the item of interest.

Example 1

- Consider, to start with, the somewhat ideal situation where all users in the database have rated the item of interest.
- In this case, $\mathcal{R}_n = \{1, \ldots, n\}$, and the asymptotic conditions on k_n become $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.
- These are just the well-known conditions ensuring consistency of the usual NN regression estimate.
- Györfi, Kohler, Krzyżak and Walk, 2002.

Example 1

- Consider, to start with, the somewhat ideal situation where all users in the database have rated the item of interest.
- In this case, $\mathcal{R}_n = \{1, \ldots, n\}$, and the asymptotic conditions on k_n become $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.
- These are just the well-known conditions ensuring consistency of the usual NN regression estimate.
- Györfi, Kohler, Krzyżak and Walk, 2002.

Example 1

- Consider, to start with, the somewhat ideal situation where all users in the database have rated the item of interest.
- In this case, $\mathcal{R}_n = \{1, \ldots, n\}$, and the asymptotic conditions on k_n become $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.
- These are just the well-known conditions ensuring consistency of the usual NN regression estimate.
- Györfi, Kohler, Krzyżak and Walk, 2002.

Example 1

- Consider, to start with, the somewhat ideal situation where all users in the database have rated the item of interest.
- In this case, $\mathcal{R}_n = \{1, \ldots, n\}$, and the asymptotic conditions on k_n become $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.
- These are just the well-known conditions ensuring consistency of the usual NN regression estimate.
- `Györfi, Kohler, Krzyżak and Walk, 2002`.

Example 2

- Fix, for simplicity, $\mathcal{R}_1 = \{1\}$.
- At step $n \geq 2$, we decide or not to add one element to $\mathcal{R}_{n-1}$ with probability $p \in (0,1)$, independently of the data.
  1. $\mathcal{R}_n$ is increased by picking a random variable B_n uniformly over the set $\{1, \ldots, n\} - \mathcal{R}_{n-1}$, and set $\mathcal{R}_n = \mathcal{R}_{n-1} \cup \{B_n\}$.
  2. Otherwise, $\mathcal{R}_n = \mathcal{R}_{n-1}$.
- Clearly, $|\mathcal{R}_n| - 1$ has binomial $\mathcal{B}(n-1, p)$ distribution.
- Consequently,

$$\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] = \frac{k_n\left[1-(1-p)^n\right]}{np}.$$

- In this setting, consistency holds provided $k_n \to \infty$ and $k_n = o(n)$ as $n \to \infty$.

Example 2

- Fix, for simplicity, $\mathcal{R}_1 = \{1\}$.
- At step $n \geq 2$, we decide or not to add one element to $\mathcal{R}_{n-1}$ with probability $p \in (0, 1)$, independently of the data.
  1. $\mathcal{R}_n$ is increased by picking a random variable B_n uniformly over the set $\{1, \ldots, n\} - \mathcal{R}_{n-1}$, and set $\mathcal{R}_n = \mathcal{R}_{n-1} \cup \{B_n\}$.
  2. Otherwise, $\mathcal{R}_n = \mathcal{R}_{n-1}$.
- Clearly, $|\mathcal{R}_n| - 1$ has binomial $\mathcal{B}(n-1, p)$ distribution.
- Consequently,

$$\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] = \frac{k_n\left[1-(1-p)^n\right]}{np}.$$

- In this setting, consistency holds provided $k_n \to \infty$ and $k_n = o(n)$ as $n \to \infty$.

Example 2

- Fix, for simplicity, $\mathcal{R}_1 = \{1\}$.
- At step $n \geq 2$, we decide or not to add one element to $\mathcal{R}_{n-1}$ with probability $p \in (0, 1)$, independently of the data.
  1. $\mathcal{R}_n$ is increased by picking a random variable B_n uniformly over the set $\{1, \ldots, n\} - \mathcal{R}_{n-1}$, and set $\mathcal{R}_n = \mathcal{R}_{n-1} \cup \{B_n\}$.
  2. Otherwise, $\mathcal{R}_n = \mathcal{R}_{n-1}$.
- Clearly, $|\mathcal{R}_n| - 1$ has binomial $\mathcal{B}(n-1, p)$ distribution.
- Consequently,

$$\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] = \frac{k_n\left[1-(1-p)^n\right]}{np}.$$

- In this setting, consistency holds provided $k_n \to \infty$ and $k_n = \mathrm{o}(n)$ as $n \to \infty$.

Example 2

- Fix, for simplicity, $\mathcal{R}_1 = \{1\}$.
- At step $n \geq 2$, we decide or not to add one element to $\mathcal{R}_{n-1}$ with probability $p \in (0,1)$, independently of the data.
    1. $\mathcal{R}_n$ is increased by picking a random variable B_n uniformly over the set $\{1, \ldots, n\} - \mathcal{R}_{n-1}$, and set $\mathcal{R}_n = \mathcal{R}_{n-1} \cup \{B_n\}$.
    2. Otherwise, $\mathcal{R}_n = \mathcal{R}_{n-1}$.
- Clearly, $|\mathcal{R}_n| - 1$ has binomial $\mathcal{B}(n-1, p)$ distribution.
- Consequently,

$$\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] = \frac{k_n\left[1-(1-p)^n\right]}{np}.$$

- In this setting, consistency holds provided $k_n \to \infty$ and $k_n = \mathrm{o}(n)$ as $n \to \infty$.

Example 2

- Fix, for simplicity, $\mathcal{R}_1 = \{1\}$.
- At step $n \geq 2$, we decide or not to add one element to $\mathcal{R}_{n-1}$ with probability $p \in (0,1)$, independently of the data.
  1. $\mathcal{R}_n$ is increased by picking a random variable B_n uniformly over the set $\{1, \ldots, n\} - \mathcal{R}_{n-1}$, and set $\mathcal{R}_n = \mathcal{R}_{n-1} \cup \{B_n\}$.
  2. Otherwise, $\mathcal{R}_n = \mathcal{R}_{n-1}$.
- Clearly, $|\mathcal{R}_n| - 1$ has binomial $\mathcal{B}(n-1, p)$ distribution.
- Consequently,

$$\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] = \frac{k_n\left[1-(1-p)^n\right]}{np}.$$

- In this setting, consistency holds provided $k_n \to \infty$ and $k_n = o(n)$ as $n \to \infty$.

Example 2

- Fix, for simplicity, $\mathcal{R}_1 = \{1\}$.
- At step $n \geq 2$, we decide or not to add one element to $\mathcal{R}_{n-1}$ with probability $p \in (0,1)$, independently of the data.
    1. $\mathcal{R}_n$ is increased by picking a random variable B_n uniformly over the set $\{1, \ldots, n\} - \mathcal{R}_{n-1}$, and set $\mathcal{R}_n = \mathcal{R}_{n-1} \cup \{B_n\}$.
    2. Otherwise, $\mathcal{R}_n = \mathcal{R}_{n-1}$.
- Clearly, $|\mathcal{R}_n| - 1$ has binomial $\mathcal{B}(n-1, p)$ distribution.
- Consequently,

$$\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] = \frac{k_n\left[1-(1-p)^n\right]}{np}.$$

- In this setting, consistency holds provided $k_n \to \infty$ and $k_n = o(n)$ as $n \to \infty$.

Example 2

- Fix, for simplicity, $\mathcal{R}_1 = \{1\}$.
- At step $n \geq 2$, we decide or not to add one element to $\mathcal{R}_{n-1}$ with probability $p \in (0,1)$, independently of the data.
    1. $\mathcal{R}_n$ is increased by picking a random variable B_n uniformly over the set $\{1, \ldots, n\} - \mathcal{R}_{n-1}$, and set $\mathcal{R}_n = \mathcal{R}_{n-1} \cup \{B_n\}$.
    2. Otherwise, $\mathcal{R}_n = \mathcal{R}_{n-1}$.
- Clearly, $|\mathcal{R}_n| - 1$ has binomial $\mathcal{B}(n-1, p)$ distribution.
- Consequently,

$$\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] = \frac{k_n\left[1-(1-p)^n\right]}{np}.$$

- In this setting, consistency holds provided $k_n \to \infty$ and $k_n = \mathrm{o}(n)$ as $n \to \infty$.

Fact 1

Recall that, for a fixed $i \in \mathcal{R}_n$, the random variable $\mathbf{X}_i^\star = (X_{i1}^\star, \ldots, X_{id}^\star)$ is defined by

$$X_{ij}^\star = \begin{cases} X_{ij} & \text{if } j \in M \\ 0 & \text{otherwise,} \end{cases}$$

and $\mathbf{X}_i^{(n)} = \mathbf{X}_i^\star$ as soon as $M \subset M_i^{n+1-i}$. For each $i \in \mathcal{R}_n$,

$$S(\mathbf{X}^\star, \mathbf{X}_i^\star) = \cos(\mathbf{X}^\star, \mathbf{X}_i^\star) = 1 - \frac{1}{2}\,\mathrm{d}^2\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}, \frac{\mathbf{X}_i^\star}{\|\mathbf{X}_i^\star\|}\right),$$

where d is the usual Euclidean distance on $\mathbb{R}^d$.

Sketch of proof

Fact 1

Recall that, for a fixed $i \in \mathcal{R}_n$, the random variable $\mathbf{X}_i^\star = (X_{i1}^\star, \ldots, X_{id}^\star)$ is defined by

$$X_{ij}^\star = \begin{cases} X_{ij} & \text{if } j \in M \\ 0 & \text{otherwise,} \end{cases}$$

and $\mathbf{X}_i^{(n)} = \mathbf{X}_i^\star$ as soon as $M \subset M_i^{n+1-i}$. For each $i \in \mathcal{R}_n$,

$$S(\mathbf{X}^\star, \mathbf{X}_i^\star) = \cos(\mathbf{X}^\star, \mathbf{X}_i^\star) = 1 - \frac{1}{2}\,\mathrm{d}^2\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}, \frac{\mathbf{X}_i^\star}{\|\mathbf{X}_i^\star\|}\right),$$

where d is the usual Euclidean distance on $\mathbb{R}^d$.

Sketch of proof

Fact 2

Let, for all $i \geq 1$,

$$T_i = \min(k \geq i : M_i^{k+1-i} \supset M).$$

Set

$$\mathcal{L}_n = \{i \in \mathcal{R}_n : T_i \leq n\},$$

and define, for $i \in \mathcal{L}_n$,

$$W_{ni}^{\star}(\mathbf{x}^{\star}) = \begin{cases} 1/k_n & \text{if } \mathbf{X}_i^{\star} \text{ is among the } k_n\text{-MS of } \mathbf{x}^{\star} \text{ in } \{\mathbf{X}_i^{\star}, i \in \mathcal{L}_n\} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$W_{ni}^{\star}(\mathbf{x}^{\star}) = \begin{cases} 1/k_n & \text{if } \frac{\mathbf{X}_i^{\star}}{\|\mathbf{X}_i^{\star}\|} \text{ is among the } k_n\text{-NN of } \frac{\mathbf{x}^{\star}}{\|\mathbf{x}^{\star}\|} \text{ in } \left\{\frac{\mathbf{X}_i^{\star}}{\|\mathbf{X}_i^{\star}\|}, i \in \mathcal{L}_n\right\} \\ 0 & \text{otherwise.} \end{cases}$$

Sketch of proof

Fact 2

Let, for all $i \geq 1$,

$$T_i = \min(k \geq i : M_i^{k+1-i} \supset M).$$

Set

$$\mathcal{L}_n = \{i \in \mathcal{R}_n : T_i \leq n\},$$

and define, for $i \in \mathcal{L}_n$,

$$W_{ni}^{\star}(\mathbf{x}^{\star}) = \begin{cases} 1/k_n & \text{if } \mathbf{X}_i^{\star} \text{ is among the } k_n\text{-MS of } \mathbf{x}^{\star} \text{ in } \{\mathbf{X}_i^{\star}, i \in \mathcal{L}_n\} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$W_{ni}^{\star}(\mathbf{x}^{\star}) = \begin{cases} 1/k_n & \text{if } \frac{\mathbf{X}_i^{\star}}{\|\mathbf{X}_i^{\star}\|} \text{ is among the } k_n\text{-NN of } \frac{\mathbf{x}^{\star}}{\|\mathbf{x}^{\star}\|} \text{ in } \left\{\frac{\mathbf{X}_i^{\star}}{\|\mathbf{X}_i^{\star}\|}, i \in \mathcal{L}_n\right\} \\ 0 & \text{otherwise.} \end{cases}$$

Sketch of proof

Fact 2

Let, for all $i \geq 1$,

$$T_i = \min(k \geq i : M_i^{k+1-i} \supset M).$$

Set

$$\mathcal{L}_n = \{i \in \mathcal{R}_n : T_i \leq n\},$$

and define, for $i \in \mathcal{L}_n$,

$$W_{ni}^\star(\mathbf{x}^\star) = \begin{cases} 1/k_n & \text{if } \mathbf{X}_i^\star \text{ is among the } k_n\text{-MS of } \mathbf{x}^\star \text{ in } \{\mathbf{X}_i^\star, i \in \mathcal{L}_n\} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$W_{ni}^\star(\mathbf{x}^\star) = \begin{cases} 1/k_n & \text{if } \frac{\mathbf{X}_i^\star}{\|\mathbf{X}_i^\star\|} \text{ is among the } k_n\text{-NN of } \frac{\mathbf{x}^\star}{\|\mathbf{x}^\star\|} \text{ in } \left\{\frac{\mathbf{X}_i^\star}{\|\mathbf{X}_i^\star\|}, i \in \mathcal{L}_n\right\} \\ 0 & \text{otherwise.} \end{cases}$$

Sketch of proof

- We write

$$\mathbb{E}\left|\sum_{i\in\mathcal{R}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}-\varphi(\mathbf{X}^\star)\right|$$
$$\le \mathbb{E}\left[\sum_{i\in\mathcal{L}_n^c} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}\right]+\mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}-\varphi(\mathbf{X}^\star)\right|.$$

- Next,

$$\mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}-\varphi(\mathbf{X}^\star)\right| \le s\mathbb{P}(\mathcal{A}_n)+\mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}^\star(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^\star\|}-\varphi(\mathbf{X}^\star)\right|,$$

where

$$\mathcal{A}_n=\left[\exists i\in\mathcal{L}_n^c : \mathbf{X}_i^{(n)} \text{ is among the } k_n\text{-MS of } \mathbf{X}^\star \text{ in } \{\mathbf{X}_i^{(n)}, i\in\mathcal{R}_n\}\right].$$

Sketch of proof

- We write

$$\mathbb{E}\left|\sum_{i\in\mathcal{R}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|} - \varphi(\mathbf{X}^\star)\right|$$
$$\leq \mathbb{E}\left[\sum_{i\in\mathcal{L}_n^c} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}\right] + \mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|} - \varphi(\mathbf{X}^\star)\right|.$$

- Next,

$$\mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|} - \varphi(\mathbf{X}^\star)\right| \leq s\mathbb{P}(\mathcal{A}_n)+\mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}^\star(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^\star\|} - \varphi(\mathbf{X}^\star)\right|,$$

where

$$\mathcal{A}_n = \left[\exists i \in \mathcal{L}_n^c : \mathbf{X}_i^{(n)} \text{ is among the } k_n\text{-MS of } \mathbf{X}^\star \text{ in } \{\mathbf{X}_i^{(n)}, i \in \mathcal{R}_n\}\right].$$

Regularity assumption

The function

$$\varphi(\mathbf{x}^\star) = \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|}\middle|\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} = \frac{\mathbf{x}^\star}{\|\mathbf{x}^\star\|}\right]$$

satisfies a Lipschitz-type property with respect to the similarity $\bar{S}$. That is, there exists a constant $C > 0$ such that, for all $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbb{R}^d$,

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{x}')| \leq C\sqrt{1 - \bar{S}(\mathbf{x}, \mathbf{x}')}.$$

In particular, for $\mathbf{x}$ and $\mathbf{x}' \in \mathbb{R}^d - \mathbf{0}$ with the same null components, this property can be rewritten as

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{x}')| \leq \frac{C}{\sqrt{2}}\,\mathrm{d}\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{x}'}{\|\mathbf{x}'\|}\right).$$

Regularity assumption

The function

$$\varphi(\mathbf{x}^\star) = \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|}\middle|\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} = \frac{\mathbf{x}^\star}{\|\mathbf{x}^\star\|}\right]$$

satisfies a Lipschitz-type property with respect to the similarity $\bar{S}$. That is, there exists a constant $C > 0$ such that, for all $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbb{R}^d$,

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{x}')| \leq C\sqrt{1 - \bar{S}(\mathbf{x}, \mathbf{x}')}.$$

In particular, for $\mathbf{x}$ and $\mathbf{x}' \in \mathbb{R}^d - \mathbf{0}$ with the same null components, this property can be rewritten as

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{x}')| \leq \frac{C}{\sqrt{2}}\,\mathrm{d}\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{x}'}{\|\mathbf{x}'\|}\right).$$

Rates of convergence

Theorem

Let $\alpha_{ni} = \mathbb{P}(M^{n+1-i} \not\ni M \,|\, M)$. *Then there exists* $C > 0$ *such that, for all* $n \geq 1$,

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| \leq C \left\{ \mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\sum_{i \in \mathcal{R}_n} \mathbb{E}\,\alpha_{ni}\right] + \mathbb{E}\left[\prod_{i \in \mathcal{R}_n} \alpha_{ni}\right] + \mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{P_n}\right] + \frac{1}{\sqrt{k_n}} \right\},$$

where $P_n = 1/(|M| - 1)$ *if* $k_n \leq |\mathcal{R}_n|$, *and* $P_n = 1$ *otherwise.*

- The performance improves as the α_{ni} decrease, i.e., as the proportion of rated items growths.
- The term $\mathbb{E}[(k_n/|\mathcal{R}_n|)^{P_n}]$ can be interpreted as a bias term in dimension $|M| - 1$, whereas $1/\sqrt{k_n}$ represents a variance term.

Rates of convergence

Theorem

Let $\alpha_{ni} = \mathbb{P}(M^{n+1-i} \not\supset M \,|\, M)$*. Then there exists* $C > 0$ *such that, for all* $n \geq 1$,

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| \leq C \left\{ \mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\sum_{i\in\mathcal{R}_n}\mathbb{E}\,\alpha_{ni}\right] + \mathbb{E}\left[\prod_{i\in\mathcal{R}_n}\alpha_{ni}\right] + \mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{P_n}\right] + \frac{1}{\sqrt{k_n}} \right\},$$

where $P_n = 1/(|M|-1)$ *if* $k_n \leq |\mathcal{R}_n|$*, and* $P_n = 1$ *otherwise.*

- The performance improves as the α_{ni} decrease, i.e., as the proportion of rated items growths.
- The term $\mathbb{E}[(k_n/|\mathcal{R}_n|)^{P_n}]$ can be interpreted as a bias term in dimension $|M|-1$, whereas $1/\sqrt{k_n}$ represents a variance term.

Rates of convergence

Theorem

Let $\alpha_{ni} = \mathbb{P}(M^{n+1-i} \not\ni M \,|\, M)$*. Then there exists* $C > 0$ *such that, for all* $n \geq 1$,

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| \\ \leq C\left\{\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\sum_{i\in\mathcal{R}_n}\mathbb{E}\,\alpha_{ni}\right] + \mathbb{E}\left[\prod_{i\in\mathcal{R}_n}\alpha_{ni}\right] + \mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{P_n}\right] + \frac{1}{\sqrt{k_n}}\right\},$$

where $P_n = 1/(|M|-1)$ *if* $k_n \leq |\mathcal{R}_n|$*, and* $P_n = 1$ *otherwise.*

- The performance improves as the α_{ni} decrease, i.e., as the proportion of rated items growths.
- The term $\mathbb{E}[(k_n/|\mathcal{R}_n|)^{P_n}]$ can be interpreted as a bias term in dimension $|M|-1$, whereas $1/\sqrt{k_n}$ represents a variance term.

Example 1

- In this ideal model, $\mathcal{R}_n = \{1, \ldots, n\}$.
- Suppose in addition that $M = \{1, \ldots, d\}$.
- Then the upper bound becomes

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)\right| = \mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/(d-1)} + \frac{1}{\sqrt{k_n}}\right).$$

- There is no influence of the dynamical rating process.
- In particular, the choice $k_n \sim n^{2/(d+1)}$ leads to

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)\right| = \mathrm{O}\left(n^{-1/(d+1)}\right).$$

Example 1

- In this ideal model, $\mathcal{R}_n = \{1, \ldots, n\}$.
- Suppose in addition that $M = \{1, \ldots, d\}$.
- Then the upper bound becomes

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)\right| = \mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/(d-1)} + \frac{1}{\sqrt{k_n}}\right).$$

- There is no influence of the dynamical rating process.
- In particular, the choice $k_n \sim n^{2/(d+1)}$ leads to

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)\right| = \mathrm{O}\left(n^{-1/(d+1)}\right).$$

Example 1

- In this ideal model, $\mathcal{R}_n = \{1, \ldots, n\}$.
- Suppose in addition that $M = \{1, \ldots, d\}$.
- Then the upper bound becomes

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| = \mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/(d-1)} + \frac{1}{\sqrt{k_n}}\right).$$

- There is no influence of the dynamical rating process.
- In particular, the choice $k_n \sim n^{2/(d+1)}$ leads to

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| = \mathrm{O}\left(n^{-1/(d+1)}\right).$$

Example 1

- In this ideal model, $\mathcal{R}_n = \{1, \ldots, n\}$.
- Suppose in addition that $M = \{1, \ldots, d\}$.
- Then the upper bound becomes

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| = \mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/(d-1)} + \frac{1}{\sqrt{k_n}}\right).$$

- There is no influence of the dynamical rating process.
- In particular, the choice $k_n \sim n^{2/(d+1)}$ leads to

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| = \mathrm{O}\left(n^{-1/(d+1)}\right).$$

Example 1

- In this ideal model, $\mathcal{R}_n = \{1, \ldots, n\}$.
- Suppose in addition that $M = \{1, \ldots, d\}$.
- Then the upper bound becomes

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)\right| = \mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/(d-1)} + \frac{1}{\sqrt{k_n}}\right).$$

- There is no influence of the dynamical rating process.
- In particular, the choice $k_n \sim n^{2/(d+1)}$ leads to

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)\right| = \mathrm{O}\left(n^{-1/(d+1)}\right).$$

Example 2

- At time 1, the user rates exactly 4 items by randomly guessing in $\{1, \ldots, d\}$.
- At time 2, he updates his preferences by adding exactly one rating among his unrated items, randomly chosen in $\{1, \ldots, d\} - M_1^1$.
- Similarly, at time 3, the user revises his preferences according to a new item uniformly selected in $\{1, \ldots, d\} - M_1^2$, and so on.
- In such a scenario, $M^j = \{1, \ldots, d\}$ for $j \geq d - 3$. Calculations show that

$$\alpha_{ni} = \begin{cases} 0 & \text{if } i \leq n - d + 4 \\ 1 - \dfrac{\binom{d-4}{n-i}}{\binom{d}{n+4-i}} & \text{if } n - d + 5 \leq i \leq n. \end{cases}$$

Example 2

- At time 1, the user rates exactly 4 items by randomly guessing in $\{1, \ldots, d\}$.
- At time 2, he updates his preferences by adding exactly one rating among his unrated items, randomly chosen in $\{1, \ldots, d\} - M_1^1$.
- Similarly, at time 3, the user revises his preferences according to a new item uniformly selected in $\{1, \ldots, d\} - M_1^2$, and so on.
- In such a scenario, $M^j = \{1, \ldots, d\}$ for $j \geq d - 3$. Calculations show that

$$\alpha_{ni} = \begin{cases} 0 & \text{if } i \leq n - d + 4 \\ 1 - \dfrac{\binom{d-4}{n-i}}{\binom{d}{n+4-i}} & \text{if } n - d + 5 \leq i \leq n. \end{cases}$$

Example 2

- At time 1, the user rates exactly 4 items by randomly guessing in $\{1, \ldots, d\}$.
- At time 2, he updates his preferences by adding exactly one rating among his unrated items, randomly chosen in $\{1, \ldots, d\} - M_1^1$.
- Similarly, at time 3, the user revises his preferences according to a new item uniformly selected in $\{1, \ldots, d\} - M_1^2$, and so on.
- In such a scenario, $M^j = \{1, \ldots, d\}$ for $j \geq d - 3$. Calculations show that

$$\alpha_{ni} = \begin{cases} 0 & \text{if } i \leq n - d + 4 \\ 1 - \dfrac{\binom{d-4}{n-i}}{\binom{d}{n+4-i}} & \text{if } n - d + 5 \leq i \leq n. \end{cases}$$

Example 2

- At time 1, the user rates exactly 4 items by randomly guessing in $\{1, \ldots, d\}$.
- At time 2, he updates his preferences by adding exactly one rating among his unrated items, randomly chosen in $\{1, \ldots, d\} - M_1^1$.
- Similarly, at time 3, the user revises his preferences according to a new item uniformly selected in $\{1, \ldots, d\} - M_1^2$, and so on.
- In such a scenario, $M^j = \{1, \ldots, d\}$ for $j \geq d - 3$. Calculations show that

$$\alpha_{ni} = \begin{cases} 0 & \text{if } i \leq n - d + 4 \\ 1 - \dfrac{\binom{d-4}{n-i}}{\binom{d}{n+4-i}} & \text{if } n - d + 5 \leq i \leq n. \end{cases}$$

Example 2

- We obtain
$$\mathbb{E}\left[\prod_{i\in\mathcal{R}_n} \alpha_{ni}\right] \leq \frac{C}{n}.$$

- Also,
$$\mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{P_n}\right] \leq C\left(\frac{k_n}{n}\right)^{1/3}.$$

- Putting all pieces together,
$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)\right| = \mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/3} + \frac{1}{\sqrt{k_n}}\right).$$

- In particular, the choice $k_n \sim n^{2/5}$ leads to
$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)\right| = \mathrm{O}(n^{-1/5}).$$

Example 2

- We obtain

$$\mathbb{E}\left[\prod_{i\in\mathcal{R}_n}\alpha_{ni}\right]\leq\frac{C}{n}.$$

- Also,

$$\mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{P_n}\right]\leq C\left(\frac{k_n}{n}\right)^{1/3}.$$

- Putting all pieces together,

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star)-\eta(\mathbf{X}^\star)\right|=\mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/3}+\frac{1}{\sqrt{k_n}}\right).$$

- In particular, the choice $k_n\sim n^{2/5}$ leads to

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star)-\eta(\mathbf{X}^\star)\right|=\mathrm{O}(n^{-1/5}).$$

Example 2

- We obtain
$$\mathbb{E}\left[\prod_{i\in\mathcal{R}_n}\alpha_{ni}\right]\leq\frac{C}{n}.$$
- Also,
$$\mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{P_n}\right]\leq C\left(\frac{k_n}{n}\right)^{1/3}.$$
- Putting all pieces together,
$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star)-\eta(\mathbf{X}^\star)\right|=\mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/3}+\frac{1}{\sqrt{k_n}}\right).$$
- In particular, the choice $k_n\sim n^{2/5}$ leads to
$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star)-\eta(\mathbf{X}^\star)\right|=\mathrm{O}(n^{-1/5}).$$

Example 2

- We obtain

$$\mathbb{E}\left[\prod_{i\in\mathcal{R}_n}\alpha_{ni}\right]\leq\frac{C}{n}.$$

- Also,

$$\mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{P_n}\right]\leq C\left(\frac{k_n}{n}\right)^{1/3}.$$

- Putting all pieces together,

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star)-\eta(\mathbf{X}^\star)\right|=\mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/3}+\frac{1}{\sqrt{k_n}}\right).$$

- In particular, the choice $k_n \sim n^{2/5}$ leads to

$$\mathbb{E}\left|\eta_n(\mathbf{X}^\star)-\eta(\mathbf{X}^\star)\right|=\mathrm{O}(n^{-1/5}).$$