M2 ISL, Sécurité Appliquée
Protection de la vie privée-PVP

Jean-François Couchot

Université de Franche-Comté, UFR-ST

Plan

Big data et vie privée

Publication non sure de données

Un premier modèle de PVP : le k-anonymat

Plan

Plan

PVP : Exacerbé par le Big Data

Big Data & Data mining

- L'exploration de données (Data Mining) : inférence de connaissances intéressantes à partir de grandes quantités de données (Big Data)
- Tendances générale : exploration des données en croissance // Big Data
- Analyse/extraction de connaissance : techniquement réalisable aujourd'hui

Big Data : application et volume

- Domaines : veille économique, découverte scientifique, santé, profilage
- Marché du Big Data en santé : $\approx$ 67,82 G$ en 2025 (Globe News Wire)
- 90% de toutes les données : créées au cours des deux dernières années (IBM)
- 97,2% des organisations : investissent en Big Data et IA. (New Vantage)
- Offres d'emploi dans le domaine : $\approx$ 2,7M en 2020. (Forbes)

Plan

Vie Privée ? [3]

Historique

- Expression du “droit d’être laissé tranquille” (the right to be let alone) [1]
- “Nul ne sera l’objet d’immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, [...]. Toute personne a droit à la protection de la loi contre de telles immixtions ou de telles atteintes.” [2]
- A l’heure d’Internet et des données (personnelles) transmises par : smartphones, messageries, GPS, appareils de fitness, moteur de recherche...

Des exemples d’inférences problématiques

- Réfrigérateur commandant des produits consommés : $\rightsquigarrow$ nbre. de présents/absents au domicile, risque sanitaire ? assurance ?
- Application de suivi de la santé : positions, fréquences card. partagées avec Apple/Google seulement ?

1. Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. Harvard law review, 193-220.
2. Organisation des Nations Unies, (1949), Déclaration universelle des droits de l’homme.
3. https ://openclassrooms.com/fr/courses/5280946-protegez-les-donnees-personnelles

Scandales de non protection de la vie privée

National Security Agency (NSA) PRISM (2007-2013)

- En 2007, création par la NSA du programme PRISM de surveillance des communications échangées sur les services en ligne des GAFAM notamment
- Transmission des données brutes à des pays tiers
- Espionnage de leaders politiques internationaux (dont A.Merkel)

Cambridge Analytica (CA) (2014-2018)

- “This is your digital life” : application pour Facebook qui permet l’aspiration de données personnelles présentes du réseau
- 87M comptes utilisateurs Facebook aspirés dès 2014
- Ciblage politique pour convaincre de voter pour D. Trump en 2016
- ”Sans CA, il n’y aurait pas eu de Brexit” selon C. Wylie

Plan

Rôles des personnes accédant à la donnée - 1

Fournisseur de contenu

- Exemple d'une photo : les personnes sur la photo + photographe ev.

Autres entités connues du fournisseur avec lesquelles il souhaite la partager

- Exemple d'une photo : amis pour partager, fournisseur de service
- Moyens utilisés : application, mécanisme de groupes et de permissions
- Règlement Général sur la Protection des Données (RGPD) :
  - A. 6 : "Données minimales collectées et traitées de manière loyale et licite"
  - A. 12 : "Transparence des informations et des communications et modalités de l'exercice des droits de la personne concernée"
  - A. 16 : "Droit à la rectification"
  - A. 17 : "Droit à l'effacement, à l'oubli"
  - A. 20 : "Droit à la portabilité" : données récupérables (format lisible)

Rôles des personnes accédant à la donnée - 2

Gestionnaire de cette donnée

- Exemple d'une photo : employés de réseau social, d'entreprises de stockage dans le nuage ;
- Catégorie critique :
  - Modèle économique du service : extraction d'information
  - Administrateurs de BD : souvent honnête mais à risque
  - Chiffrement des données : souvent une solution

Reste du monde

- Exemple d'une photo : celles/ceux avec lequel on ne veut pas la partager

Plan

Plan

BD statistiques : introduction

Motivations pieuse

- Permettre l'accès à des statistiques concernant des groupes d'individus
- Tout en restreignant l'accès aux informations individuelles

Pb. : des vestiges d'informations individuelles contenus dans chq. stat.

- Comparaison des salaires totaux de deux groupes qui ne diffèrent que d'un individu
- Possibilité de déduire le salaire de l'individu manquant à l'un

BD statistiques : modèle [4]

Les M attributs d'une relation (issue d'une jointure ev.)

- Chaque attribut A_j, $1 \leq j \leq M$: $|A_j| \geq 2$ valeurs possibles
- x_{ij} : valeur de l'attribut A_j pour l'individu i
- Hypothèse simplificatrice : un seul enregistrement par individu
- N enregistrements

Caractérisation de sous-groupe par une formule C

- C : formule en logique propositionnelle
- Opérateurs : "+" (OR), "." (AND), "——" (NOT)
- Relation issue de C : ensemble des enregistrements qui vérifient C

4. Denning, D. E. R. (1982). Cryptography and data security (Vol. 112). Reading : Addison-Wesley.

BD statistique : exemple de base

La base

Nom	Sexe	Dpt.	Année	Test	Note
Allen	F	Info	00	600	3,4
Baker	F	Ing	00	520	2,5
Cook	H	Ing	98	630	3,5
Davis	F	Info	98	800	4,0
Evans	H	Bio	99	500	2,2
Frank	H	Ing	01	580	3,0
Good	H	Info	98	700	3,8
Hall	F	Psy	99	580	2,8
Ilies	H	Info	01	600	3,2
Jones	F	Bio	99	750	3,8
Kline	F	Psy	01	500	2,5
Lane	H	Ing	98	600	3,0
Moore	H	Info	99	650	3,5

- Identifiant : Nom
- Champs sensibles : Note, Test
- Exemple de sous groupe C : $(\text{Sexe} = H).((\text{Dpt} = \text{Info}) + (\text{Dpt} = \text{Bio}))$, en abrégé en $H.(\text{Info} + \text{Bio})$
- Relation issue de C : $\{(\text{Evans}, H, \text{Bio}, 99, 500, 2, 2), (\text{Good}, H, \text{Info}, 98, 700, 3, 8), (\text{Ilies}, H, \text{Info}, 01, 600, 3, 2), (\text{Moore}, H, \text{Info}, 99, 650, 3, 5)\}$

BD statistiques : requêtes

Statistiques simples

- Informations uniquement agrégées par
  - $\texttt{count}(C) = |C|$
  - $\texttt{sum}(C, A_j) = \sum_{i \in C} x_{ij}$
- Exemple : $\texttt{count}(H.(Info + Bio)) = |H.(Info + Bio)| = 4$
- Autres requêtes identifiantes interdites : ~~$F.\text{Info} = \{\text{Allen, Davis}\}$~~

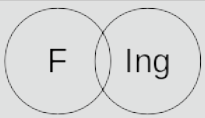
Statistiques sensibles

- Statistique déclarée comme sensible : si dévoile trop d'informations confidentielles à propos d'un individu
- Statistiques interdites : celles issues d'un singleton
- Exemple $F.\text{Ing}$ est le singleton $\{(\text{Baker}, F, \text{Ing}, 00, 520, 2.5)\}$
  $\texttt{sum}(F.\text{Ing}, \text{Note})$ est sensible et sera interdite
- Idée : pour un seuil n choisi, n'autoriser que les requêtes sur des relations de cardinalité dans $[n, N - n]$,

Requêtes statistiques dans $[n, N-n]$: pas sures

Tracker individuel, ordre 2

Nom	Sexe	Dpt.	Année	Test	Note
Baker	F	Ing	00	520	2,5
Cook	H	Ing	98	630	3,5
Frank	H	Ing	01	580	3,0
Lane	H	Ing	98	600	3,0



- $|F.\text{Ing}| = 1$: sensible !
- $|\text{Ing}| = 4$, $[\overline{F}.\text{Ing}| = 3$: OK
- $\text{sum}(F.\text{Ing}, \text{Note}) = \text{sum}(\text{Ing}, \text{Note}) - \text{sum}(\overline{F}.\text{Ing}, \text{Note}) = 12,0 - 9,5 = 2,5$

Tracker individuel, ordre 3

Nom	Sexe	Dpt.	Année	Test	Note
Evans	H	Bio	99	500	2,2
Hall	F	Psy	99	580	2,8
Jones	F	Bio	99	750	3,8
Moore	H	Info	99	650	3,5

99 \ H.Bio	00	01	11	10
0				
1			X	

- $q(H.\text{Bio}.99) + q((\overline{H} + \overline{\text{Bio}}).99) = q(99)$
- $q(H.\text{Bio}.99) = q(99) - q((\overline{H} + \overline{\text{Bio}}).99)$
- $\text{sum}(H.\text{Bio}.99, \text{Note}) = \text{sum}(99, \text{Note}) - \text{sum}((\overline{H} + \overline{\text{Bio}}).99, \text{Note}) = 12,3 - 10,1 = 2,2$

Inconvénient : nécessite de caractériser l'individu

Requêtes statistiques dans $[n, N-n]$: pas sures

Tracker général T : partager l'espace en deux

	T	$\overline{T}$
C	w	x
$\overline{C}$	y	z

- $q(\text{Tous}) = q(T) + q(\overline{T}) = w + x + y + z$
- Essayer $q(C) = q(C+T) + q(C+\overline{T}) - q(\text{Tous})$
- Sinon $q(C) = 2 \times q(\text{Tous}) - q(\overline{C}+T) - q(\overline{C}+\overline{T})$

Jones possède F.Bio, sans test d'unicité

- Essai avec le tracker $T = H$.
- $|\text{Tous}| = |H| + |\overline{H}| = 7 + 6 = 13$
- $|F.\text{Bio}| = |F.\text{Bio} + H| + |F.\text{Bio} + \overline{H}| - |\text{Tous}| = 1\,!$
- $sum(\text{Tous}, \text{Note}) = sum(H, \text{Note}) + sum(\overline{H}, \text{Note}) = 41,2$
- $sum(F.\text{Bio}, \text{Note}) =$
  $sum(F.\text{Bio} + H, \text{Note}) + sum(F.\text{Bio} + \overline{H}, \text{Note}) - sum(\text{Tous}, \text{Note}) = 3,8$

Présentation agrégée : pas toujours sure

Macrostatistiques

- Collections de statistiques agrégées, usuellement présentées sous la forme de tableaux à double entrées
- Inconvénient : réduction extrême de l'utilité des données, difficulté d'établir des corrélations entre attributs...

Exemple de macrostatistiques

Effectifs par année/sexe

Sexe	98	99	00	01	Total
F	1	2	2	1	6
H	3	2	0	2	7
Total	4	4	2	3	13

Tests par année/sexe

Sexe	98	99	00	01	Total
F	800	1330	1120	500	3750
H	30	1150	0	1120	4260
Total	2730	2480	0	1180	8010

Suppressions

Tests par année/sexe

Sexe	98	99	00	01	Total
F	*	1330	1120	*	3750
H	1930	1150	0	1120	4260
Total	2730	2480	0	1180	8010

Tests par année/sexe

Sexe	98	99	00	01	Total
F	*	1330	1120	*	3750
H	*	1150	0	*	4260
Total	2730	2480	0	1180	8010

Pas de garantie de sûreté.

Arrondi systématique de valeurs : pas sûr

Arrondi systématique de q dans une base b

- Soit $b' = \left\lfloor \frac{b+1}{2} \right\rfloor$ et $d \equiv q \mod (b)$
- $r(q) = \begin{cases} q & \text{si } d = 0 \\ q - d & \text{si } d < b' \\ q + (b - d) & \text{si } d \geq b' \end{cases}$
- $r(q) \in [r(q) - b' + 1, r(q) + b' - 1]$

	$r(q_i)$	l_i	u_i
q_1	15	13	17
q_2	10	8	12
q_3	15	13	17
q_4	20	18	22
$\sum$		52	**68**
Q	70	$L = \mathbf{68}$	$U = 72$

Cas d'arrondis réversibles pour un attribut A

- Ensembles disjoints $C_1, \ldots, C_m$ et $q_i = sum(C_i, A)$, $1 \leq i \leq m$,
- $Q = q_1 + \cdots + q_m$ le total de toutes les sommes
- $[l_i, u_i]$ l'intervalle estimant $r(q_i)$, $[L, U]$ l'intervalle estimant $r(Q)$
- Si $L = \sum_i u_i \rightarrow \forall i, 1 \leq i \leq m, q_i = u_i$
- Si $U = \sum_i l_i \rightarrow \forall i, 1 \leq i \leq m, q_i = l_i$

Plan

Pseudonymisation

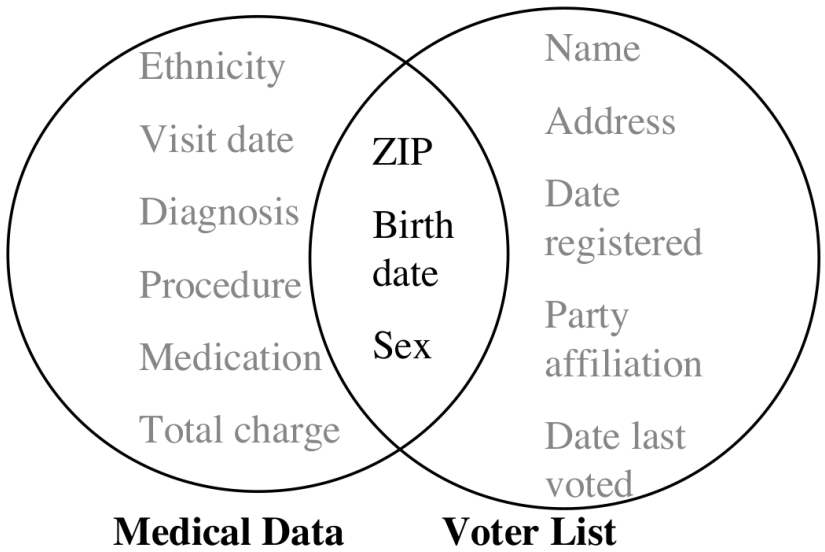
- Champs identifiants : supprimés et remplacés par un id (H(NSS)).

	Non-sensibles				Sensibles
H	CP	Age	Genre	Nationalité	Pathologie
c4ca4238a0b923820dcc509a6f75849b	13053	28	H	russe	trouble cardiaque
c81e728d9d4c2f636f067f89cc14862c	13068	29	H	américaine	trouble cardiaque
eccbc87e4b5ce2fe28308fd9f2a7baf3	13068	21	F	japonaise	infection virale
a87ff679a2f3e71d9181a67b7542122c	13053	23	H	américaine	infection virale
e4da3b7fbbce2345d7772b0674a318d5	14853	49	H	indienne	cancer
1679091c5a880faf6fb5e6087eb1b2dc	14853	48	F	russe	trouble cardiaque
8f14e45fceea167a5a36dedd4bea2543	14850	47	H	américaine	infection virale
c9f0f895fb98ab9159f51fd0297e236d	14850	49	F	américaine	infection virale
45c48cce2e2d7fbdea1afc51c7c6ad26	13053	31	H	américaine	cancer
d3d9446802a44259755d38e6d163e820	13053	37	H	indienne	cancer
6512bd43d9caa6e02c990b0a82652dca	13068	36	F	japonaise	cancer
c20ad4d76fe97759aa27a0c99bff6710	13068	35	F	américaine	cancer

- Avantage : calculs identiques à ceux sur la base de données initiale (Age moyen/cancer=37,8)

Pseudonymisation : attaque par quasi-identifiants [5]

- Base de donnée médicale pseudonymisée et publique



- Liste d'électeurs publique, recensement USA, 1990 : "87% of the population in the US had characteristics that likely made them unique based only on 5-digit Zip, gender, date of birth"
- CP, genre, date de naissance : quasi-identifiants
- Identification de données médicales du gouverneur Weld

5. Sweeney, L. (2002). k-anonymity : A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), 557-570.

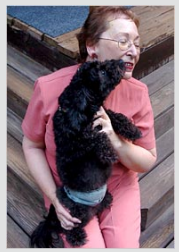
Pseudonymisation : attaque par intersection [6]

2006, diffusion par AOL de 20M requêtes, 658K utilisateur sans nom

AnonID	Query	QueryTime
1326	“holiday mansion houseboat”	2006-03-29
1326	“back to the future”	2006-04-01
591476	“english spanish translator”	2006-03-20
591476	“panama vacations”	2006-03-20
591476	“breast reduction”	2006-03-23
591476	“volunteer work at hospitals in brooklyn”	2006-05-24
591476	...	...
591476	“how to secretly poison your ex”	2006-03-12

Thelma Arnold, 62 ans, veuve vivant à Lilburn, Ga., réidentifiée en 3 j.

AnonID	Query
4417749	“people with last name ’Arnold”’
4417749	“landscapers in Lilburn, Ga”
4417749	“60 single men”
4417749	“dog that urinates on everything”
4417749	dog-related queries



⇝ Suppression hâtive des données sur le site d’AOL.

6. BARBARO, Michael, ZELLER, Tom, et HANSELL, Saul. A face is exposed for AOL searcher no. 4417749. New York Times, 2006, vol. 9, no 2008, p. 8.

Plan

Plan

Le k-anonymat[5] et les quasi-identifiants -1

Quasi-identifiants : QID

- QID, intuition[7] : "des éléments qui ne sont pas en eux-mêmes des identificateurs uniques, mais qui sont suffisamment bien corrélés avec une entité pour pouvoir être combinés avec d'autres quasi-identifiants afin de créer un identificateur unique"
- QID, définition[8] : Les attributs de $Q \subseteq \{A_1, \ldots, A_M\}$ sont quasi-identifiants de la relation T si la requête suivante retourne au moins un résultat

```
SELECT Q FROM T
GROUP BY Q
HAVING COUNT(*)=1
```

- (CP, genre, date de naissance) : triplets uniques dans 87% des cas $\rightsquigarrow$ quasi-identifiants

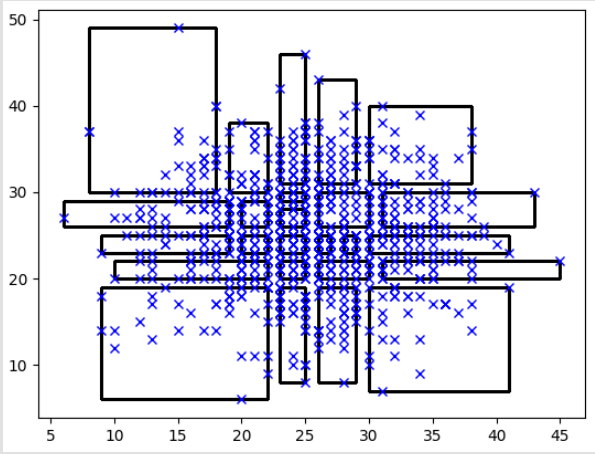
7. https://en.wikipedia.org/wiki/Quasi-identifier

8. Nguyen, B., & Castelluccia, C. (2020). Techniques d'anonymisation tabulaire : concepts et mise en oeuvre. arXiv preprint arXiv :2001.02650.

Le k-anonymat[5] et les quasi-identifiants -2

Intuition : regrouper les QID pour casser l'unicité

- Niveau de détail des valeurs des QID : à réduire pour qu'il y ait au moins k individus différents dont les QIDs sont égaux
- Individus avec mêmes QIDs : font partie de la même classe d'équivalence



Plan

k-anonymat par généralisation

Réduction des niveaux de détail

Sur l'exemple :

- CP : laisser, **regroupements par 2**, suppr.
- Age : laisser, par intervalles d'amplitudes 10, 20,**suppr.**
- Genre : laisser, **suppr.**
- Nationalité : laisser, par continent, **suppr.**
- $\rightsquigarrow 3 \times 4 \times 2 \times 3 = 72$ combinaisons de généralisation !

Regroupement par classes d'équivalence de card. $\geq$ à 4

CP	Age	Genre	Nationalité	Pathologie
	[20; 30[	*	*	trouble cardiaque
{13053	[20; 30[	*	*	trouble cardiaque
13058}	[20; 30[	*	*	infection virale
	[20; 30[	*	*	infection virale
	[40; 50[	*	*	cancer
{14850	[40; 50[	*	*	trouble cardiaque
14853}	[40; 50[	*	*	infection virale
	[40; 50[	*	*	infection virale
	[30; 40[	*	*	cancer
{13053	[30; 40[	*	*	cancer
13058}	[30; 40[	*	*	cancer
	[30; 40[	*	*	cancer

} 4 individus

} 4 individus

} 4 individus

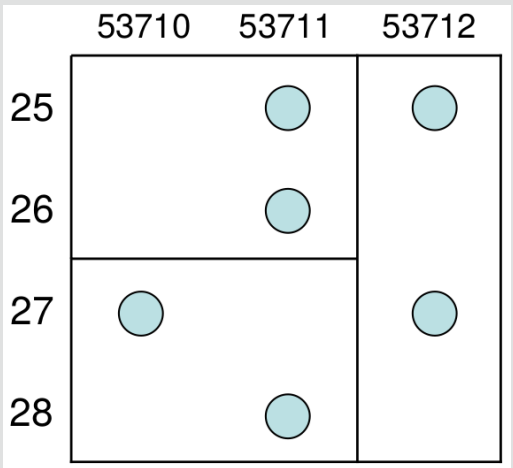
k-anonymat par Mondrian [9]

Algorithme général, en factorielle

1. Pour chaque séq. $[A_1, \ldots, A_f]$ de QID, telle qu'il reste plus de k éléments par groupe
   - partitionner intelligemment selon la médiane des données de A_i
2. généraliser les valeurs des attributs : un groupe $\equiv$ une généralisation
3. évaluer la perte d'information

2-anonymat, partitionnement selon [CP, Age, Genre]

Age	Genre	CP	Pathologie
25	H	53711	Grippe
25	F	53712	Hépatite
26	H	53711	Bronchite
27	H	53710	Bras cassé
27	F	53712	SIDA
28	H	53711	Ongle perdu



Age	Genre	CP	Pathologie
[25-26]	H	53711	Grippe
[25-27]	F	53712	Hépatite
[25-26]	H	53711	Bronchite
[27-28]	H	[53710-53711]	Bras cassé
[25-27]	F	53712	SIDA
[27-28]	H	[53710-53711]	Ongle perdu

9. LeFevre, K., DeWitt, D. J., & Ramakrishnan, R. (2006, April). Mondrian multidimensional k-anonymity. In 22nd International conference on data engineering (ICDE'06) (pp. 25-25). IEEE.

C_{AVG} : nbre. moy. d'élements par classe normalisé

Définition pour $|T|$ enregistrements et propriétés

$$C_{AVG} = \frac{|T|}{|EQs|} \times \frac{1}{k}$$

- et $|EQs|$: nbre. de classes d'équiv.
- $\frac{|T|}{|EQs|}$ nbre. moy. d'éléments par classe $(\geq k) \rightsquigarrow C_{AVG} \geq 1$
- Utilité décroissante à mesure que C_{AVG} croît.

Exemple avec 4-anonymat

CP	Age	Genre	Nationalité	Pathologie
	[20; 30[	*	*	trouble cardiaque
{13053	[20; 30[	*	*	trouble cardiaque
13058}	[20; 30[	*	*	infection virale
	[20; 30[	*	*	infection virale
	[40; 50[	*	*	cancer
{14850	[40; 50[	*	*	trouble cardiaque
14853}	[40; 50[	*	*	infection virale
	[40; 50[	*	*	infection virale
	[30; 40[	*	*	cancer
{13053	[30; 40[	*	*	cancer
13058}	[30; 40[	*	*	cancer
	[30; 40[	*	*	cancer

- $C_{AVG} = \frac{12}{3} \times \frac{1}{4} = 1$
- Optimal pour cette métrique

Loss : Perte dûe à la généralisation

Définition pour $|T|$ enregistrements, n QID et propriétés

$$Loss = \frac{1}{n|T|} \sum_{j=1}^{|T|} \sum_{i=1}^{n} \frac{R_{i_j}}{R_i}$$

- R_{i_j}/R_i : rapport acquis/acquérables
  - discret : $(|\text{généralisation } i_j| - 1)/(|\text{Qid } i| - 1)$
  - continu : $(U_{i_j} - L_{i_j})/(U_i - L_i)$
- Moyenne des acquis/acquérables, utilité décroissante // *Loss* croît

Exemple avec 4-anonymat

CP	Age	Genre	Nationalité	Pathologie
{13053 13058}	[20; 30[	*	*	trouble cardiaque
	[20; 30[	*	*	trouble cardiaque
	[20; 30[	*	*	infection virale
	[20; 30[	*	*	infection virale
{14850 14853}	[40; 50[	*	*	cancer
	[40; 50[	*	*	trouble cardiaque
	[40; 50[	*	*	infection virale
	[40; 50[	*	*	infection virale
{13053 13058}	[30; 40[	*	*	cancer
	[30; 40[	*	*	cancer
	[30; 40[	*	*	cancer
	[30; 40[	*	*	cancer

- Nat. $= \{\text{russe}, \text{am.}, \text{jap.}, \text{ind.}\} \rightsquigarrow$ $R_{\text{Nat}} = 3$, $R_{\text{CP}} = 3$, $R_{\text{Genre}} = 1$,
- $R_{\text{Age}} = 50 - 20 = 30$
- $Loss = \frac{1}{4 \times 12} \left(12 \times \left(\frac{1}{3} + \frac{10}{30} + \frac{1}{1} + \frac{3}{3}\right)\right) = \frac{2}{3}$

Divergence de Kullback-Leibler

Mesure de dissimilarité entre deux distributions

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log_2 \left(\frac{P(x)}{Q(x)} \right).$$

- P (resp. Q) distrib. bassée sur les données réelles (resp. modifiées)
- Positive, utilité décroissante avec l'accroisement de D_{KL}

Exemple avec 4-anonymat

CP	Age	Genre	Nationalité	Pathologie
{13053 13058}	[20; 30[	*	*	trouble cardiaque
	[20; 30[	*	*	trouble cardiaque
	[20; 30[	*	*	infection virale
	[20; 30[	*	*	infection virale
{14850 14853}	[40; 50[	*	*	cancer
	[40; 50[	*	*	trouble cardiaque
	[40; 50[	*	*	infection virale
	[40; 50[	*	*	infection virale
{13053 13058}	[30; 40[	*	*	cancer
	[30; 40[	*	*	cancer
	[30; 40[	*	*	cancer
	[30; 40[	*	*	cancer

- $P(13053, 28, H, \textit{russe}) = \cdots = P(13068, 35, F, \textit{américaine}) = \frac{1}{12}$
- $Q(13053, 28, H, \textit{russe}) = \cdots = Q(13068, 35, F, \textit{américaine}) = \frac{1}{3} \times \left(1 - (1 - \frac{1}{2 \times 10 \times 2 \times 4})^4\right)$
- $D_{\mathrm{KL}}(P \parallel Q) = 12 \times \frac{1}{12} \log_2(\frac{\frac{1}{12}}{Q(13053,28,H,\textit{russe})}) \approx 3,35$

k-anonymat : attaques

Exemple

CP	Age	Genre	Nationalité	Pathologie
	[20; 30[	*	*	trouble cardiaque
{13053	[20; 30[	*	*	trouble cardiaque
13058}	[20; 30[	*	*	infection virale
	[20; 30[	*	*	infection virale
	[40; 50[	*	*	cancer
{14850	[40; 50[	*	*	trouble cardiaque
14853}	[40; 50[	*	*	infection virale
	[40; 50[	*	*	infection virale
	[30; 40[	*	*	cancer
{13053	[30; 40[	*	*	cancer
13058}	[30; 40[	*	*	cancer
	[30; 40[	*	*	cancer

4 individus

4 individus

4 individus

Attaques

- Homogénéité :
  - $\oplus$ Patient de 35 ans connu $\rightsquigarrow$ cancer.
  - $\ominus$ Patient de 29 ans connu $\rightsquigarrow$ ~~cancer~~.
- Connaissance supplémentaire : un japonais de 21 ans, P(trouble cardiaque|japonais)=faible $\rightsquigarrow$ infection virale.