La méthode des clés aléatoires (*Cell Key Method*)

Julien Jamme[1]

***Résumé*— L'objectif de cette note méthodologique est de décrire de façon rapide la méthode des clés aléatoires, méthode perturbatrice qui permet d'assurer une diffusion de données agrégées respectant la confidentialité. La note présente les principes de la méthode, son fonctionnement illustré sur un cas simple et un premier exemple d'application par l'Insee.**

I. La protection des données tabulées

L'essentiel des données statistiques diffusées par le service statistique public (SSP) est diffusé sous forme agrégée et plus particulièrement de tableaux. Même si elles ne contiennent pas de données personnelles directement identifiantes, ces données sont susceptibles de conduire indirectement à des ré-identifications ou à des divulgations d'informations confidentielles. C'est le cas en particulier lorsque les cases d'un tableau représentent seulement quelques individus (ex. case en rouge de la table I). Plusieurs méthodes existent pour se prémunir contre ces risques de divulgation.

TABLE I – Table originale contenant une case sensible (seuil arbitraire)

	CAT 1	CAT 2	CAT 3	**Ensemble**
A	4	5	10	**19**
B	12	20	7	**39**
Ensemble	**16**	**25**	**17**	**58**

A. *Méthodes suppressives*

Classiquement, l'Insee traite ce genre de situation problématique (un exemple est fourni en table I) de deux manières différentes et complémentaires :

— Soit en reformant le tableau initial afin de ne plus avoir de cases contenant des petits comptages, c'est-à-dire en réarrangeant les catégories d'une variable, par exemple en fusionnant les catégories 1 et 2 de la table I (voir table II).

TABLE II – Solution 1 : Table après regroupement des catégories 1 et 2

	CAT 1 et 2	CAT 3	**Ensemble**
A	9	10	**19**
B	32	7	**39**
Ensemble	**41**	**17**	**58**

— Soit en supprimant l'information contenue dans les cases (on appelle cela la suppression primaire). Cette solution nécessite néanmoins de supprimer d'autres cases afin d'empêcher de reconstruire l'information confidentielle masquée (on appelle cela la suppression secondaire). Pour notre exemple, la suppression minimisant la valeur des cases cachées conduit à supprimer les cases internes des catégories 1 et 3 (table III, suppression primaire en rouge et suppressions secondaires en orange).

[1]J. Jamme, Insee, Département des Méthodes Statistiques, dg75-dmcsi-confidentialite@insee.fr

TABLE III – Solution 2 : Table après suppression des informations

	CAT 1	CAT 2	CAT 3	**Ensemble**
A		5		**19**
B		20		**39**
Ensemble	**16**	**25**	**17**	**58**

Les méthodes suppressives conduisent ainsi à réduire l'information qui est diffusée et la présence de cases blanchies rend difficile la manipulation des tableaux par l'utilisateur (par exemple s'il souhaite agréger des informations). En revanche, la donnée disponible est restée intacte.

B. *Méthodes perturbatrices*

Une autre approche consiste à perturber les informations diffusées en injectant un peu de bruit dans les données, comme dans la table IV.

Le premier intérêt d'une telle approche est de rendre possible la diffusion d'une information dans toutes les cases du tableau. L'intensité de la perturbation est choisie de telle sorte que la confidentialité soit préservée tout en limitant son impact sur les données, celles-ci conservant leur utilité. En revanche, la perturbation peut engendrer une perte de l'additivité des catégories (voir par exemple la première colonne de la table IV).

Une revue des méthodes perturbatrices applicables à des données tabulées est présentée dans le chapitre 5 de [Hundepool *et al.*, 2024]. Nous présentons ici la méthode des clés aléatoires - ou *cell key method* en anglais - car elle présente de nombreux avantages par rapport à toutes ses concurrentes.

TABLE IV – Solution 3 : Table après une légère perturbation des informations

	CAT 1	CAT 2	CAT 3	**Ensemble**
A	6	3	9	**20**
B	12	18	8	**39**
Ensemble	**19**	**24**	**18**	**59**

II. Principes de la méthode des clés aléatoires

La méthode des clés aléatoires a été développée initialement par [Fraser et Wooton, 2005] pour l'Australian Bureau of Statistics (ABS) en 2005. Elle est présentée de façon plus complète dans [Thompson *et al.*, 2013] et des développements complémentaires ont été réalisés par des statisticiens de l'institut fédéral allemand de statistique (Destatis) dans [Enderle *et al.*, 2018] et [Gießing et Tent, 2019]. Elle a fait partie des méthodes de gestion de la confidentialité préconisées par Eurostat pour la diffusion des données du recensement européen de 2021.

La méthode a d'abord été conçue pour être appliquée sur des données de comptages. Nous nous concentrerons sur cet aspect par la suite.

La méthode des clés aléatoires consiste à injecter du bruit dans *toutes* les cases des tableaux qui sont diffusés à partir d'une même source. Elle se distingue d'autres méthodes de perturbation sur deux points essentiels :

1) Le bruit injecté dans une case d'un tableau dépend des individus qui la composent, grâce à l'attribution d'un **jeu de clés individuelles tirées aléatoirement** au préalable.
2) Le mécanisme de perturbation est entièrement piloté par **une matrice de transition** décrivant les distributions de probabilité des perturbations à injecter.

De ces deux points découlent les caractéristiques suivantes :

— **Toutes les cases** d'un tableau agrégé seront perturbées, ce qui assure une protection contre les différenciations en tout genre et notamment géographiques.
— Les cases sont **perturbées indépendamment** les unes des autres.
— La méthode des clés aléatoires est dotée de la propriété : « les mêmes contributeurs, la même perturbation » : une case qui se trouverait dans des tableaux différents serait toujours perturbée de la même manière.
— **Les zéros originaux ne sont pas perturbés**, ceci permet de ne pas générer des situations non réalistes (comme des enfants à la retraite).

La méthode nécessite de fixer la valeur maximale D et la variance maximale V du bruit ajouté au comptage initial.

A. Notations

Pour la suite de la présentation, on utilisera les notations suivantes :

— N, le nombre d'individus d'une base de données individuelles $\mathscr{B}$.
— $X \in [\![0;N]\!]$, la variable aléatoire représentant la valeur originale d'une case $\mathscr{C}$ d'un tableau. Il s'agit donc du nombre d'individus ayant les caractéristiques décrites par la case en question.
— $Z \in [\![-D;+D]\!]$, la variable aléatoire représentant le bruit à injecter.
— $X' \in \mathbb{N}$, la variable aléatoire représentant la valeur de la même case après perturbation, ainsi X' est définie par $X' = X + Z$.
— $D \in \mathbb{N}^*$, la déviation maximale.
— $V > 0$ la variance du bruit Z.
— $p_{ij} = \mathbb{P}(X' = j | X = i)$, la probabilité de transition de $X = i$ vers $X' = j$.
— $\mathscr{M}$, le mécanisme de bruitage transformant X en X'.

Pour empêcher l'apparition de valeurs négatives, on distingue deux cas :

$$\begin{cases} Z \in [\![-D;+D]\!] \text{ si } X \geq D \\ \text{ou} \\ Z \in [\![-X;+D]\!] \text{ sinon} \end{cases}$$

B. Caractéristiques du mécanisme

Dans la méthode des clés aléatoires, le mécanisme de bruitage $\mathscr{M}$ est piloté par une matrice de transition $\mathbf{M}$ dont chaque ligne décrit la distribution de probabilités de la perturbation d'une valeur originale. Ainsi pour $X = i$ fixé, on a $\forall j \in [\![0;N+D]\!]$, $\mathbf{M}[i,j] = p_{ij}$.

Définir le mécanisme de bruitage consiste donc à déterminer les probabilités de transition en fonction des paramètres D et V souhaités et de façon à respecter les propriétés suivantes :

— Si $i = 0, \forall\ j\ p_{ij} = 0$: les zéros ne sont pas bruités.
— Si $i \in [\![0;D-1]\!]$, alors le support de la loi sera $j \in [\![0;i+D]\!]$.
— Si $i \geq D$, alors le support de la loi sera $j \in [\![i-D;i+D]\!]$. Dans ce cas, on appliquera d'ailleurs la même distribution de probabilités quelque soit la valeur i de départ.
— La matrice $\mathbf{M}$ sera donc de taille $(D+1) \times (2D+1)$.
— Le bruit injecté sera **sans biais** et de variance maximale V.

Par exemple, pour $D = 2$, la matrice $\mathbf{M}$ aura la forme suivante :

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ p_{10} & p_{11} & p_{12} & p_{13} & 0 \\ p_{20} & p_{21} & p_{22} & p_{23} & p_{24} \end{bmatrix}$$

En pratique, la matrice de transition est générée avec le package `R ptable` ([Enderle, 2023]) par maximisation de l'entropie sous les contraintes énoncées plus haut.

C. Les clés aléatoires

La méthode tient son nom du fait qu'**elle associe une clé aléatoire à chaque case d'un tableau agrégé**. Cette clé - nommée *cell key* en anglais - est construite de la manière suivante :

1) Une **clé individuelle**, notée rk_l, est d'abord associée à chaque individu l de la base $\mathscr{B}$. Il s'agit d'un nombre décimal aléatoire tiré selon une loi uniforme entre 0 et 1.
2) La clé associée à la case $\mathscr{C}$ d'un tableau agrégé est la **partie décimale de la somme des clés des individus la composant**. Pour une case $\mathscr{C}$, la clé associée $ck_{\mathscr{C}}$ est définie par :

$$ck_{\mathscr{C}} = \sum_{l \in \mathscr{C}} rk_l - \lfloor \sum_{l \in \mathscr{C}} rk_l \rfloor$$

où $\lfloor x \rfloor$ désigne la partie entière (inférieure) de x.

Par construction, la clé des cases du tableau suit également une loi uniforme entre 0 et 1. Il s'agit donc bien d'un nombre aléatoire complètement indépendant du tableau dans lequel elle apparaît : seule la composition de la case influe sur sa valeur.

D. *La table de perturbation*

Pour appliquer la perturbation à partir de la matrice de transition **M** et des clés aléatoires des cases du tableau, il faut se reporter à une table dite *table de perturbation*, dont un exemple est fourni ci-dessous (table V). Les quatre premières colonnes sont directement issues de la matrice de transition. Les deux dernières correspondent aux intervalles des probabilités p_{ij} cumulées.

TABLE V – Table de perturbation pour $D = 2$ et $V = 1$

$X = i$	$X' = j$	p_{ij}	$Z = z$	p_{inf}	p_{sup}
0	0	1	0	0	1
1	0	0.366	-1	0	0.366
1	1	0.366	0	0.366	0.733
1	2	0.168	$+1$	0.733	0.901
1	3	0.099	$+2$	0.901	1
$i \geq 2$	$i-2$	0.064	-2	0	0.064
$i \geq 2$	$i-1$	0.245	-1	0.064	0.309
$i \geq 2$	i	0.383	0	0.309	0.691
$i \geq 2$	$i+1$	0.245	$+1$	0.691	0.936
$i \geq 2$	$i+2$	0.064	$+2$	0.936	1

Ainsi, si $X = 1$ et que la clé de la case vaut 0.75, alors on appliquera une déviation de $+1$ (ligne de la table V surlignée en bleu, correspondant à l'emplacement de la flèche sur la figure 1). On diffuserait alors la valeur $X' = 2$.

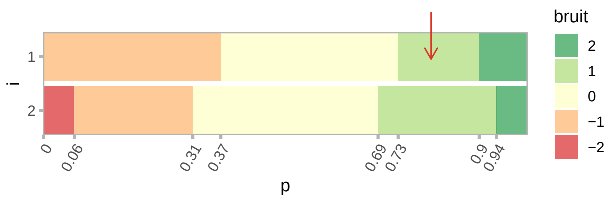

FIG. 1 – Version graphique de la table de perturbation V

III. UN EXEMPLE ÉTAPE PAR ÉTAPE

Pour illustrer, le plus simplement possible, le processus qui conduit à perturber des données tabulées, on partira d'une base constituée uniquement de 6 individus (table VI). Imaginons que l'on souhaite diffuser deux tableaux : le nombre d'habitants par commune de résidence et le nombre d'habitants par âge. On choisit $D = 2$ et $V = 1$.

A. *Étape 1 : Tirage des clés individuelles*

La première étape consiste à **tirer les clés aléatoires individuelles** pour les associer à chaque ligne (individu) de notre jeu de données (dernière colonne de la base présentée VI).

TABLE VI – Ajout des clés

id	Commune de résidence	Âge	Clé
1	Amiens	25	0,9177275
2	Paris	20	0,8850062
3	Marseille	45	0,6266963
4	Amiens	45	0,1117820
5	Marseille	20	0,6496634
6	Marseille	20	0,2813433

B. *Étape 2 : Construction du tableau*

La deuxième étape consiste à **construire le tableau agrégé** - ici le nombre d'individus par commune de résidence (table VII). En même temps que le dénombrement des individus (colonne 3), on agrège les clés aléatoires individuelles (colonne 4) dont on ne conserve que la partie décimale pour obtenir une clé aléatoire associée à chaque case du tableau (colonne 5).

TABLE VII – Tableau agrégé par Commune de résidence et calcul de la clé associé à la case

Com.	ids	X	$\sum$ clés	Clé de la case
Amiens	{1,4}	2	1,0295095	0,0295095
Marseille	{3,5,6}	3	1,5577030	0,5577030
Paris	{2}	1	0,8850062	0,8850062
Total	{1,2,3,4,5,6}	6	3,4722187	0,4722187

C. *Table de perturbation*

La troisième étape consiste à **construire une table de perturbation** qui permettra d'indiquer quel bruit injecter selon la clé et la valeur de la case. La table V présentée plus haut correspond exactement au jeu de paramètres fixés pour cet exemple ($D = 2$ et $V = 1$). Cette table peut être facilement construite avec le package R `ptable` en utilisant le code suivant :

```
ptable::create_cnt_ptable(D=2, V=1)@pTable
```

D. Rapprocher le tableau agrégé et la table de perturbation

La quatrième et dernière étape consiste à **rapprocher le tableau agrégé (table VII) à la table de perturbation (table V)**. Pour la commune d'Amiens, le processus est le suivant :

— Son effectif original vaut 2 : on repère les lignes de la table de perturbation VIII dont la première colonne $X = i$ prend la valeur 2 (ici, les cinq dernières).

— La clé de la case vaut $0,0295095$ (table VII) : on repère parmi les lignes sélectionnées l'intervalle $[p_{inf}, p_{sup}[$ tel que $0,0295095$ soit inclus dedans (ici l'intervalle correspondant est $[0, 0.064[$). Une seule ligne doit correspondre (ici, celle surlignée en vert dans la table VIII).

— On applique alors le bruit correspondant (4^e colonne) à la ligne repérée pour perturber l'effectif d'Amiens : $X' = X + Z = 2 - 2 = 0$.

TABLE VIII – Détection dans la table de perturbation du bruit à ajouter pour la commune d'Amiens

$X = i$	$X' = j$	p_{ij}	$Z = z$	p_{inf}	p_{sup}
0	0	1	0	0	1
1	0	0.366	-1	0	0.366
1	1	0.366	0	0.366	0.733
1	2	0.168	$+1$	0.733	0.901
1	3	0.099	$+2$	0.901	1
$i \geq 2$	$i-2$	0.064	-2	0	0.064
$i \geq 2$	$i-1$	0.245	-1	0.064	0.309
$i \geq 2$	i	0.383	0	0.309	0.691
$i \geq 2$	$i+1$	0.245	$+1$	0.691	0.936
$i \geq 2$	$i+2$	0.064	$+2$	0.936	1

Le processus est appliqué de la même manière pour obtenir l'effectif perturbé de chaque commune (dernière colonne de la table IX).

TABLE IX – La table bruitée

Com.	Clé de la case	X	Z	X'
Amiens	0,0295095	2	-2	0
Marseille	0,5577030	3	0	3
Paris	0,8850062	1	+1	2
Total	0,4722187	6	0	6

En calculant les effectifs par âge et en leur appliquant le processus décrit plus haut, on obtient la perturbation présentée dans la table X, dans lequel on observera que le total est perturbé exactement de la même manière que dans la table IX.

TABLE X – Effectifs par âge avant et après perturbation

Âge	ids	Clé de la case	X	Z	X'
20	{2,5,6}	0,8160129	3	+1	4
25	{1}	0,9177275	1	+2	3
45	{3,4}	0,7384783	2	+1	3
Total	{1,2,3,4,5,6}	0,4722187	6	0	6

IV. Arbitrage risque-utilité

Le choix des paramètres de la méthode - principalement [1] la déviation D et la variance V - constitue un enjeu important puisqu'il permettra d'assurer le meilleur compromis entre une protection optimale de la confidentialité des données et la meilleure conservation possible de leur utilité.

A. Utilité des données

La méthode des clés aléatoires permet d'obtenir un meilleur compromis risque-utilité que des méthodes d'arrondis par exemple. Et globalement, toutes les analyses statistiques reposant sur les tableaux diffusés reproduisent très bien les analyses qui auraient été réalisées sur les données originales. Néanmoins, la perturbation ayant relativement plus d'effet sur les petits comptages que sur les grands, il est conseillé de ne pas interpréter des statistiques qui reposeraient sur de trop petits comptages ($\leq 2 \times D$).

En revanche, à cause de la perturbation systématique, l'utilisateur ne retrouvera pas l'une des caractéristiques habituelles des données tabulées : l'additivité (Comme signalé dans l'exemple ci-dessus dans les tables IX et X). Pour l'utilisation des totaux, deux préconisations peuvent être faites :

— Si le but est d'interpréter le comptage lui-même, il est préférable d'utiliser la valeur perturbée du total, si elle est disponible, plutôt que la somme des modalités perturbées.

— En revanche, si le but est de construire des tableaux de % ventilant les différentes modalités, il est préférable d'utiliser la somme des modalités perturbées au dénominateur. Ceci permet à la fois d'obtenir une ventilation additive et des ratios légèrement plus précis.

B. Influence de la variance

La variance V est le paramètre qui influence le plus le compromis risque-utilité. Les trois graphiques de la figure 2 présentent les distributions de probabilités où, pour $D = 2$, la variance vaut respectivement 0.5 (vert), 1 (orange) et 10 (violet).

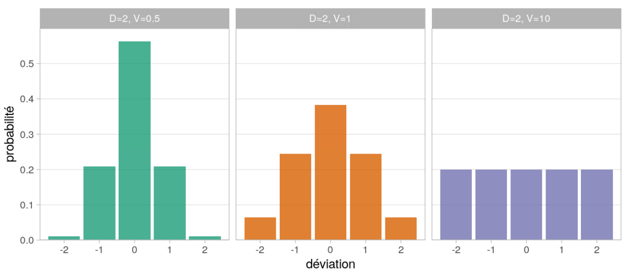

FIG. 2 – Distributions de probabilités décrivant trois scénarios d'injection de bruit différents

1. Il est aussi possible d'interdire l'apparition de certaines valeurs après perturbation. On interdira par exemple la présence de petits comptages tels que les 1 et les 2 dans les tableaux diffusés afin de renforcer la protection des données.

Chacun de ces graphiques décrit donc une façon différente de perturber les données :

- **Scénario 1 (vert)** : privilégie la préservation de l'information, en modifiant peu les données. En effet, 56% des cases ne seront pas modifiées et seulement 2% des cases seront déviées de ± 2 (table XI). Un bruitage aussi faible protège malheureusement moins bien les données.
- **Scénario 3 (violet)** : privilégie la protection des données en perturbant fortement les données. En effet, seulement 20% des cases ne seront pas déviées et 40% subiront une déviation maximale (de ± 2). Il conduira à une perte plus conséquente de l'information comparativement au premier scénario.
- **Scénario 2 (orange)** : scénario intermédiaire. Il perturbe davantage l'information que le scénario 1 et conduit à mieux protéger l'information mais sans entraîner une perte d'information aussi importante que le scénario 3.

TABLE XI – Caractéristiques des trois scénarios d'injection de bruit présentés dans la figure 2

	Scénario 1 (Vert)	**Scénario 2 (Orange)**	**Scénario 3 (Violet)**
Variance	0.5	1	10
% de chances que X ne soit pas dévié	56%	38%	20%
% de chances que X subisse une déviation maximale (± 2)	2%	13%	40%
Conservation de l'information	++	+	–
Niveau de protection	–	+	++

Ce petit exemple illustre le besoin de choisir des paramètres qui assurent le meilleur compromis entre protection contre le risque de divulgation et diffusion d'une information de qualité.

C. Mesurer le risque de divulgation avec les probabilités de transition inverses

Une fois les données perturbées, il est intéressant de pouvoir mesurer la capacité d'un utilisateur/attaquant à inférer le vrai comptage à partir des informations dont il dispose.

À la suite de [Enderle *et al.*, 2018], on utilise les probabilités $q_{ij} = \mathbf{P}(X = i|X' = j)$, qu'on appellera *probabilités de transition inverses*. En calculant ces probabilités, on pourra ainsi estimer la capacité d'inférence d'un attaquant en fonction des paramètres retenus pour définir le mécanisme et ainsi s'assurer que la protection est suffisante.

Ces probabilités peuvent s'exprimer à partir des probabilités de transition p_{ij} grâce à la formule de Bayes et à la formule des probabilités totales :

$$q_{ij} = \mathbf{P}(X = i|X' = j) = \frac{p_{ij} \times \mathbf{P}(X = i)}{\sum_{k \in \mathbb{N}} p_{kj} \times \mathbf{P}(X = k)} \quad (1)$$

Dans l'équation 1, les probabilités de transition p_{ij}, p_{kj} sont connues. En revanche, les probabilités $\mathbf{P}(X = i)$ ne peuvent être qu'estimées, soit à partir des données réelles pour le producteur, soit à partir des seules données perturbées pour l'utilisateur/attaquant.

Le producteur peut estimer ces probabilités *ex ante* à partir des fréquences observées et ainsi mesurer la capacité d'inférence d'un attaquant, qui servira de mesure objective du risque de divulgation de données confidentielles.

D. La protection contre la différenciation géographique

La différenciation géographique consiste à retrouver une information non diffusée directement en utilisant des informations disponibles sur des zonages géographiques qui se superposent, comme l'illustre la figure 3. Une information y est diffusée à la fois au niveau communal et sur des carreaux. Avec un seuil de confidentialité fixé à 5, aucune information mise à disposition ne serait confidentielle dans ce cas. Néanmoins, en retirant à la population de la commune A la somme de l'ensemble des carreaux inclus entièrement dans celle-ci, nous obtenons une information confidentielle sur la partie dans A du carreau intersectant les deux communes : $100 - (15 + 59 + 7 + 16) = 3$ (figure 4).

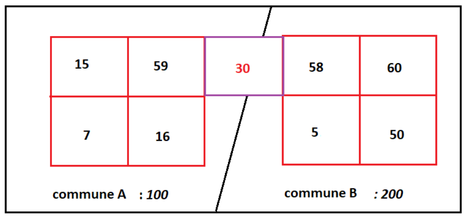

FIG. 3 – Informations diffusées

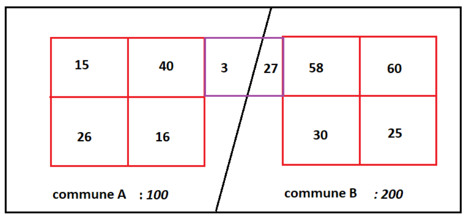

FIG. 4 – Informations divulguées

Mais en perturbant légèrement l'ensemble des informations diffusées, c'est-à-dire en ajoutant une déviation aléatoire comprise entre $-D$ et $+D$ à chaque comptage, on rend la différenciation incertaine :

$$\begin{aligned}&(100 + d1) - ((15 + d2) + (59 + d3) + (7 + d4) + (16 + d5))\\&= 3 + (d1 - (d2 + d3 + d4 + d5))\end{aligned}$$

où $d1, d2, d3, d4, d5 \in [\![-D; +D]\!]$.

Chaque information diffusée étant bruitée, le résultat de la différenciation l'est nécessairement plus, rendant l'incertitude plus forte : dans notre exemple, le support du bruit associé à la valeur différenciée est $\llbracket -5 \cdot D ; +5 \cdot D \rrbracket$ et sa variance n'est pas V mais $5 \cdot V$. La méthode des clés aléatoires, comme toute méthode perturbatrice intégrale, permet donc de réduire très largement ce risque de différenciation géographique. Cette protection est obtenue à moindre coût pour le producteur, ce qui constitue un avantage certain pour le producteur.

V. Application à l'Insee

À l'occasion de l'actualisation des contours des Quartiers prioritaires de la politique de la ville (QPV) qui a eu lieu en 2024, l'Insee applique cette méthode sur un grand nombre d'indicateurs diffusés sur les QPV, comme le nombre de demandeurs d'emplois, d'allocataires de la CAF ou de logements du parc social locatif. Les indicateurs portant sur les revenus ou issus du recensement de la population font l'objet d'un traitement différent. La méthode des clés aléatoires vient remplacer des méthodes de blanchiment (suppressives) utilisées jusqu'alors.

Ce changement est d'abord motivé par le fait que l'Insee souhaite continuer à diffuser ces mêmes données sur les anciens contours des QPV qui datent de 2015. Cela permet de répondre à un besoin exprimé par les utilisateurs principaux de ces données. Or, les deux contours pouvant être très proches, cette double diffusion engendre de très nombreux risques de différenciation géographique, rendant très complexes les traitements historiques de ce type de risque. Une méthode perturbatrice permet, en revanche, de faciliter le traitement de ces cas et la protection contre ce type de risque.

La méthode des clés aléatoires a été choisie pour sa simplicité d'implémentation, et parce qu'elle est une des rares méthodes aléatoires qui assure qu'un indicateur sera toujours perturbé de la même manière quelque soit l'endroit (le tableau) où il apparaît et le moment où il a été calculé. Elle est donc tout à fait adaptée à une production décentralisée et asynchrone des tableaux statistiques, dès lors que chacun des producteurs de tableaux utilise les mêmes clés individuelles et les mêmes paramètres.

References

[Enderle, 2023] Enderle, T. (2023). *ptable : Generation of Perturbation Tables for the Cell-Key Method.* R package version 1.0.0.

[Enderle *et al.*, 2018] Enderle, T., Giessing, S. et Tent, R. (2018). *Designing Confidentiality on the Fly Methodology – Three Aspects*, volume 11126 de *Lecture Notes in Computer Science*, page 28–42. Springer International Publishing, Cham.

[Fraser et Wooton, 2005] Fraser, B. et Wooton, J. (2005). A proposed method for confidentialising tabular output to protect against differencing. *In Monographs of Official Statistics : Work Session on Statistical Data Confidentiality*, page 299–302.

[Gießing et Tent, 2019] Giessing, S. et Tent, R. (2019). Concepts for generalising tools implementing the cell key method to the case of continuous variables. *In UNECE - Expert Meeting on Statistical Data Confidentiality*, the Hague.

[Hundepool *et al.*, 2024] Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Longhurst, J., Schulte Nordholt, E., Seri, G., De Wolf, P.-P., Tent, R., Młodak, A., Gussenbauer, J. et Wilak, W. (2024). *Handbook on Statistical Disclosure Control - https ://sdc-tools.github.io/HandbookSDC/.* ESSNet SDC, 2nd edition édition.

[Thompson *et al.*, 2013] Thompson, G., Broadfoot, S. et Elazar, D. (2013). Methodology for the automatic confidentialisation of statistical outputs from remote servers at the australian bureau of statistics. *In Joint UNECE/Eurostat work session on statistical data confidentiality*, Ottawa.

Insee
Mesurer pour comprendre

Département des méthodes statistiques

14 février 2025