



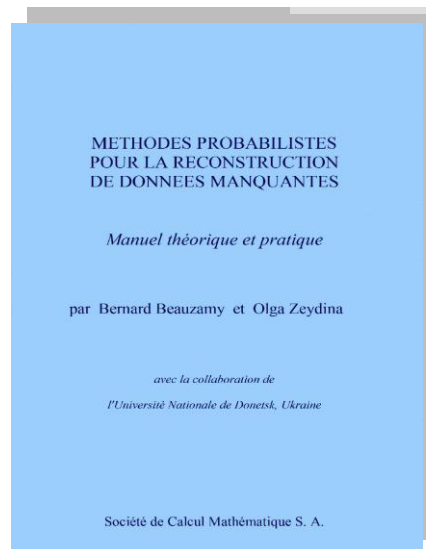
Madame, Monsieur,

Nous vous présentons le livre de

Bernard Beauzamy et Olga Zeydina :

**Méthodes Probabilistes
pour la reconstruction
de données manquantes :**

Manuel théorique et pratique



ISBN : 2-9521458-2-2, ISSN : 1767-1175. Dimensions 15,3 x 24 cm. Broché, 265 pages.

Des données manquantes, ce n'est pas cela qui manque, si l'on ose dire. Et quand ce ne sont pas les données qui manquent, ce sont les responsables qui sont en vacances et quand tout le monde est là, on constate qu'il n'y a plus d'argent. Mais ne soyons pas défaitistes dès le début !

Il peut y avoir d'excellentes raisons pour que des données soient absentes. Tout d'abord, on ne les a pas recueillies, parce que l'on ne savait pas que ce serait utile, ou bien parce que l'on ne pouvait pas les mesurer. Ensuite, on peut les avoir recueillies, mais les avoir perdues : cela arrive aux plus consciencieux. Enfin, on peut les avoir mesurées, puis recueillies, mises à l'abri, et que cependant un sort contraire les ait détruites...

Heureusement, des méthodes mathématiques existent, permettant de reconstituer les données perdues, et nous les exposons dans cet ouvrage. Il s'agit d'utiliser au mieux ce que l'on connaît pour reconstituer ce que l'on ne connaît pas : il faut donc faire appel aux probabilités. L'exemple traité illustre bien la puissance de ces méthodes : pour évaluer l'ampleur des pénuries d'eau en Vendée, à la demande de Veolia Environnement, Région Ouest, nous avons dû reconstituer les débits journaliers de 19 fleuves, enregistrés sur 37 années, mais avec environ 50 % de "trous".

Si l'on peut reconstituer des données, il n'est pas utile de les recueillir : on réalise ainsi des économies, chose bien plaisante. Nous traitons l'exemple explicite des températures de huit villes de l'Ohio : il suffirait d'en mesurer trois pour reconstituer les cinq autres au demi-degré près, économisant ainsi cinq stations de mesure !

Table des matières

Remerciements	9
Introduction	11
I. Chacun cherche ses données	11
II. Des règles contraignantes	11
III. Des exemples frappants	12
IV. Des bénéfices attendus	13
V. Ne pas promettre monts et merveilles	13
VI. Un sketch	14
V. Organisation de l'ouvrage	17
Chapitre I - Pourquoi des données manquantes ?	19
I. Principales causes	19
II. Deux situations très distinctes	21
III. Les méthodes probabilistes	22
Chapitre II - Les méthodes usuelles de reconstruction	23
I. L'Ecole déterministe	23
II. L'approche statistique	25
Chapitre III - Introduction aux méthodes probabilistes	35
I. L'horreur de l'aléatoire	35
II. Description d'un processus aléatoire	40
III. La fonction de répartition	43
IV. Perte d'information	46
V. Espérance	47
VI. La variance	49
VII. Dispersion autour de l'espérance	49
VIII. Variance et valeurs manquantes	50
IX. Unités	50
Chapitre IV - Reconstruction de données à partir de l'échantillon	53
I. Utilisation de valeurs moyennes	53
II. Utilisation des valeurs existantes voisines	56
III. La simulation	58
Chapitre V - Introduction aux probabilités conditionnelles	61
I. Notion de loi conjointe	61
II. Etablir la loi conjointe	64
III. La précision de l'information	66
IV. Répétitions indépendantes	68
V. Exploitation de la loi conjointe	68
VI. Extensions à plusieurs variables	69
Chapitre VI - Reconstitution	71
I. Reconstitution de l'Autisme à partir de la Sèvre	71
II. Deux manières de procéder	76
III. Mise en œuvre pratique	77
IV. Choix de la taille de l'intervalle	78
V. Indicateur de proximité	79
VI. Qualité de la reconstruction	83
VII. La théorie	84
VIII. Les hypothèses faites	85
Chapitre VII - Choix de l'échantillon d'appui	87
I. Généralités	87
II. Faits et méfaits du coefficient de corrélation linéaire	88
III. Coefficient de corrélation et domaines respectifs	98
Chapitre VIII - Un exemple de reconstruction	107
I. Données manquantes	107
II. Division en deux parties de l'échantillon existant	108
III. Reconstitution de la seconde partie	111
IV. Indicateur de qualité de la reconstruction	115
V. Choix de la taille de l'intervalle	116

IV. Reconstitution complète	117
V. Variances conditionnelles, reconstitution	121
VI. Précision et justesse	124
Chapitre IX - Améliorer la méthode de reconstruction	125
I. Propriétés des ensembles de conditionnement	125
II. Intervalles de taille variable	126
III. La méthode par intervalles NEVD	127
IV. Méthode sans intervalles	136
Chapitre X - La variance conditionnelle	143
I. Considérations générales	143
II. La variance conditionnelle	144
III. Indicateur préliminaire de choix	149
IV. Propriétés de la variance conditionnelle	150
V. Ajout de mesures à l'échantillon	153
VI. Dimensions et invariances	153
Chapitre XI - Travail sur l'échantillon d'appui	155
I. Décalage temporel sur l'échantillon d'appui	155
II. Combinaison de plusieurs échantillons	157
III. Conclusion	159
Chapitre XII - Utilisation de plusieurs échantillons	161
I. Utiliser toute l'information disponible	161
II. La théorie	162
III. Absence de données	166
IV. Un exemple complet	166
V. Mise en œuvre de la méthode NEVD	172
VI. Comparaison des méthodes de reconstruction	178
VII. Conclusion pratique	180
VIII. Un résultat théorique	181
Chapitre XIII - Reconstitution complète	187
I. Utilisation des variances conditionnelles	187
II. Le programme principal	190
Chapitre XIV - La qualité d'une reconstruction	193
I. Analyse de l'erreur de reconstruction	193
II. La qualité de l'appareil de mesure	197
III. Fabrication de l'appareil de mesure à distance	198
IV. Mise en œuvre de l'appareil de mesure	199
V. Limitations de l'appareil	199
VI. Caractéristiques de l'appareil de mesure	200
VII. Indicateurs globaux de qualité	203
VIII. Définition d'un indicateur des flux mensuels	210
Chapitre XV - Diminuer le nombre de mesures	213
I. Cesser de mesurer des débits en Vendée	213
II. Cesser de mesurer des températures en Ohio	220
III. Précautions à prendre	230
Chapitre XVI - Valeurs extrêmes	233
I. Observations de longue durée	233
II. Apparition de valeurs extrêmes	235
III. Les températures à Paris-Montsouris	236
IV. Comparaison à des tirages selon une loi donnée	240
V. Mise en place d'un indicateur quantitatif	244
VI. Probabilité de dépassement d'un record	246
VII. Critique des méthodes usuelles	250
VIII. Peut-on mieux faire ?	251
IX. Approche déterministe	252
X. La gestion sociale des phénomènes extrêmes	252
Bibliographie	255

Bon de commande

Je commande un/des exemplaire(s) du livre :

« Méthodes Probabilistes pour la reconstruction de données manquantes »,

Nom :

Prénom :

Société :

Adresse de livraison :

Adresse de facturation, si différente :

Montant HT : 75,83 Euros.

TVA 5,5 % : 4,17 Euros.

Montant TTC : 80 Euros.

Port inclus.

Pour les particuliers : Merci de retourner ce bon de commande, avec votre règlement ; une facture acquittée vous sera adressée.

Pour les entreprises : Merci de nous retourner ce bon de commande avec votre cachet, ou bien de nous adresser votre propre bon de commande. Nous vous adresserons la facture en même temps que le livre.

Société de Calcul Mathématique, SA, 111 Faubourg Saint Honoré, 75008 Paris
tel 01 42 89 10 89, fax 01 42 89 10 69, <http://www.scmsa.eu>