

Théorie de l'information et du codage

© LAVOISIER, 2007

LAVOISIER  
11, rue Lavoisier  
75008 Paris

[www.hermes-science.com](http://www.hermes-science.com)  
[www.lavoisier.fr](http://www.lavoisier.fr)

ISBN 978-2-7462-1719-5

---

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, d'une part, que les "copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective" et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, "toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite" (article L. 122-4). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Tous les noms de sociétés ou de produits cités dans cet ouvrage sont utilisés à des fins d'identification et sont des marques de leurs détenteurs respectifs.

---

Printed and bound in England by Antony Rowe Ltd, Chippenham, September 2007.

# Théorie de l'information et du codage

Olivier Rioul

**hermes**  
**Science**  
—publications—

*Lavoisier*



## Table des matières

<b>Introduction</b> . . . . .	9
<b>PREMIÈRE PARTIE. OUTILS DE LA THÉORIE DE L'INFORMATION</b> . . . . .	15
<b>Chapitre 1. Entropie et entropie relative</b> . . . . .	17
1.1. Rappels sur les variables aléatoires . . . . .	17
1.1.1. Variables aléatoires discrètes . . . . .	17
1.1.2. Variables aléatoires continues . . . . .	20
1.1.3. Notation unifiée . . . . .	22
1.2. Entropie $H(X)$ d'une variable aléatoire . . . . .	23
1.3. Entropie différentielle . . . . .	28
1.4. Entropie relative ou divergence $D(p, q)$ . . . . .	30
<b>Chapitre 2. Traitement et information</b> . . . . .	35
2.1. Traitements et canaux . . . . .	35
2.1.1. Traitements et canaux discrets . . . . .	37
2.1.2. Traitements et canaux continus . . . . .	42
2.1.3. Traitements et canaux réciproques . . . . .	44
2.2. Information mutuelle $I(X, Y)$ . . . . .	46
<b>Chapitre 3. Information et entropie</b> . . . . .	51
3.1. Information mutuelle et entropies . . . . .	51
3.2. Information et incertitude . . . . .	53
3.3. Diagrammes de Venn . . . . .	55
3.4. Information transmise sur des canaux discrets . . . . .	58
3.5. Information et entropies différentielles . . . . .	60
<b>Chapitre 4. Concavité et Maximum d'entropie</b> . . . . .	67

6 Théorie de l'Information et du Codage

4.1. Propriétés de concavité et de convexité . . . . .	67
4.2. Inégalité de Gibbs et borne de Shannon . . . . .	71
4.3. Inégalités de Fano . . . . .	78
<b>Chapitre 5. Chaînes de traitement et perte d'information . . . . .</b>	<b>83</b>
5.1. Chaînes de Markov . . . . .	83
5.1.1. Deux traitements successifs $X \rightarrow Y \rightarrow Z$ . . . . .	83
5.1.2. Plusieurs traitements successifs $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n$ . . . . .	88
5.2. Développement de l'information sur plusieurs v.a. . . . .	91
5.3. Traitement de données et information mutuelle . . . . .	97
5.4. Traitement de données et divergence . . . . .	102
<b>Chapitre 6. Information de Fisher et e.q.m. minimale . . . . .</b>	<b>105</b>
6.1. Information de Fisher paramétrique $J_\theta(X)$ . . . . .	105
6.2. Inégalité de Cramér-Rao . . . . .	110
6.3. Traitement de données et information de Fisher . . . . .	113
6.4. Erreur quadratique moyenne minimale $\text{Var}(\theta X)$ . . . . .	116
6.5. Traitement de données et e.q.m. minimale . . . . .	120
6.6. Information de Fisher non paramétrique et e.q.m.m. . . . .	121
<b>Chapitre 7. Variance entropique et identité de de Bruijn . . . . .</b>	<b>125</b>
7.1. Entropie et mélange de variables aléatoires . . . . .	125
7.2. Variance entropique . . . . .	130
7.3. Inégalité de l'information de Fisher . . . . .	133
7.4. Informations de Fisher et de Shannon . . . . .	137
7.5. Identité de de Bruijn . . . . .	142
<b>DEUXIÈME PARTIE. LIMITES ET THÉORÈMES DE SHANNON . . . . .</b>	<b>145</b>
<b>Chapitre 8. Sources et canaux . . . . .</b>	<b>147</b>
8.1. Modèles de sources . . . . .	147
8.2. Modèles de canaux . . . . .	153
8.3. Entropie et information mutuelle des composantes . . . . .	159
<b>Chapitre 9. Codage de source et de canal . . . . .</b>	<b>165</b>
9.1. Le problème général du codage . . . . .	165
9.2. Codage de source . . . . .	170
9.3. Codage de canal . . . . .	172
9.4. Codage de source/canal conjoint . . . . .	174
<b>Chapitre 10. Limites de Shannon . . . . .</b>	<b>177</b>
10.1. L'inégalité fondamentale du codage : OPTA . . . . .	177

10.2. Codage de source : fonction taux-distortion $R(D)$ . . . . .	180
10.3. Codage de canal : fonction capacité-coût $C(P)$ . . . . .	182
10.4. Allure des fonctions $R(D)$ et $C(P)$ . . . . .	184
10.5. Influence de la dimension . . . . .	190
10.6. Influence de la mémoire . . . . .	192
<b>Chapitre 11. Calcul théorique des limites de Shannon</b> . . . . .	197
11.1. $R(D)$ pour une source sans mémoire . . . . .	197
11.2. $C(P)$ pour un canal additif sans mémoire . . . . .	202
11.3. Divers . . . . .	207
<b>Chapitre 12. Séquences typiques</b> . . . . .	211
12.1. Séquences typiques . . . . .	211
12.2. Séquences conjointement typiques . . . . .	213
12.3. Inégalités de dépendance typique . . . . .	215
<b>Chapitre 13. Théorèmes de Shannon</b> . . . . .	217
13.1. Codage de source sans pertes . . . . .	217
13.2. Codage de canal sans contrainte de coût . . . . .	221
13.3. Codage de canal : cas général . . . . .	225
13.4. Codage de source avec pertes (cas général) . . . . .	226
13.5. Commentaires . . . . .	231
13.6. Codage source/canal . . . . .	232
13.7. L'éloge de la paresse . . . . .	235
<b>Annexes</b> . . . . .	239
A. Exercices pour la première partie . . . . .	239
B. Problèmes . . . . .	255
B.1. Codage optimal pour le canal à effacement . . . . .	255
B.2. Capacité de Hartley du canal uniforme . . . . .	256
B.3. Calcul de capacité de canaux symétriques . . . . .	257
B.4. Encadrement de la fonction taux-distorsion . . . . .	259
B.5. Encadrement de la capacité . . . . .	260
B.6. Algorithme de Blahut-Arimoto . . . . .	260
B.7. Capacité avec voie de retour . . . . .	263
B.8. Capacité d'un canal à états . . . . .	264
B.9. Capacité d'un canal à états connus . . . . .	266
B.10. Capacité du canal gaussien à évanouissements . . . . .	267
B.11. Capacité du canal de Gilbert-Elliott . . . . .	268
B.12. Entropie d'une source stationnaire . . . . .	269
B.13. Capacité d'un canal binaire avec mémoire . . . . .	270
B.14. Systèmes sûrs en cryptographie à clef secrète . . . . .	271

8 Théorie de l'Information et du Codage

B.15. Codage source-canal tandem dans le cas gaussien . . . . .	273
B.16. Région de capacité d'un canal à accès multiple . . . . .	274
<b>Bibliographie annotée . . . . .</b>	<b>279</b>
<b>Index . . . . .</b>	<b>281</b>

## Index

### A

accès multiple *voir aussi* : canal à —  
systèmes à — (TDMA, FDMA,  
CDMA) 276  
aléatoire *voir* : processus, variable —  
alphabet (de symboles) 17  
ambiguïté 52  
anormalité *voir* : non-gaussianité  
auto-information *voir* : information  
mutuelle  
auto-régressif 152, 153

### B

bande (largeur de —) 172, 205, 268, 276  
Bayes (formule de —) 44  
Bertrand (intégrale, série de —) 71  
bit (*binary unit*) 25  
Blachman *voir* : score  
Blahut-Arimoto (algorithme de —) 208,  
260–262  
blanc (bruit) 132, 136, 150, 159  
Boltzmann *voir* : entropie

### C

canal 35, 153  
à accès multiple (MAC) 274  
additif 36, 86, 140, 155, 158, 159, 202  
à états 43, 208, 264, 266  
à évanouissement 43, 267  
avec mémoire 159, 162, 195, 209, 270  
avec voie de retour 156, 208, 263

binaire à effacement (BEC) 41, 60,  
208, 241, 255  
binaire à effacement et erreur 41  
binaire symétrique (BSC) 38, 45, 58,  
158, 202, 203, 235, 247  
cauchien 141  
causal 155, 156  
de Gilbert-Elliott 208, 268  
de Rayleigh 44, 208, 267  
discret 37, 260  
discret sans mémoire (DMC) 260  
en  $\Sigma$  41  
en  $Z$  40, 59, 241, 258  
faiblement symétrique 258  
fortement symétrique 257  
gaussien (AWGN) 42, 80, 144, 159,  
204, 236  
laplacien 141, 260  
 $M$ -aire symétrique 39, 79, 203  
markovien 159  
opaque 46, 58, 102  
réciproque 44  
sans mémoire 157, 161, 193, 202, 260  
stationnaire 154  
symétrique 207, 257  
uniforme (de Hartley) 256, 260  
capacité 183  
avec voie de retour 208, 263  
du canal binaire à effacement 255, 258,  
269  
du canal binaire à effacement et erreur  
258, 269

- du canal binaire markovien 270
  - du canal binaire symétrique 202, 203
  - du canal de Gilbert-Elliott 268
  - du canal de Rayleigh 267
  - du canal gaussien 204
  - du canal  $M$ -aire symétrique 203
  - du canal uniforme 256
  - d'un canal à états 264, 266
  - d'un canal à évanouissement 267
  - région de — 274
  - capacité-coût (fonction —) 179, 182–184, 189–195, 202
  - Cauchy-Schwarz (inégalité de —) 111, 229
  - Césaro (moyenne de —) 209, 269
  - chaîne de Markov 83, 97, 150, 247
    - paramétrique 113
    - réciproque 87, 91
    - sous-chaîne 89
  - clé de cryptage 271
  - codage 165
    - aléatoire 222, 227
    - à longueur variable 219
    - à temps partagé 276
    - avec pertes 167, 171, 226
    - conjoint source/canal 174, 175, 235, 237, 273
    - de canal 172, 182, 221
    - de source 170, 180, 217, 226
    - différentiel 271
    - en tandem 174, 175, 232, 273
    - inégalité fondamentale du — *voir* : OPTA
    - sans pertes 167, 171, 181, 193, 217
    - typique 227
  - complémentarité (entre information de Fisher et erreur quadratique moyenne minimale) 123
  - compression *voir* : codage de source
  - concavité 67
    - de la variance entropique 253
    - de l'entropie 69
    - de l'information mutuelle 70, 247
    - du logarithme 32, 126, 243
    - du log-det 245
  - convexité 67
    - de la divergence 69
    - de la fonction inverse 134
    - de la fonction puissance 230
    - de l'information de Fisher 251
    - de l'information mutuelle 70, 247
  - corrélation (coefficient de —) 47, 110, 119, 152
  - coût (fonction de —) *voir* : poids
  - Cramér-Rao (inégalité de —)
    - améliorée 143
    - avec biais 249
    - matricielle 113
    - non paramétrique 122, 135, 139, 141, 253
    - paramétrique 110, 111
  - cryptage 271
    - de Vernam 272
- D**
- débit binaire 170, 172, 205
  - de Bruijn (identité de —) 142, 251, 252
  - decit (*decimal unit*) 25
  - décodage *voir aussi* : codage
    - non supervisé 265, 268
    - supervisé 265, 268
    - typique 222
  - décorrélation *voir* : corrélation
  - densité *voir* : distribution
  - dépendance
    - coefficient de — 242
    - conditionnelle 87, 92
    - et information mutuelle 47
    - typique 215
  - discrimination *voir* : divergence
  - distance *voir aussi* : divergence
    - de Hamming 158, 168, 173, 198, 199
    - de Kullback-Leibler 32
    - d'information 242
  - distorsion 167
    - voir aussi* : erreur
    - bornée 229
    - maximale 186, 188
    - minimale 186
    - quadratique 79, 168, 188, 200, 229
  - distribution (de probabilité)
    - Bêta 141
    - binomiale 243
    - conditionnelle 36, 153
    - de Bernoulli 18

de Cauchy 141  
 de Dirac 21, 22  
 de Kronecker 19, 22  
 de Pareto 240  
 de Poisson 19  
 de Rayleigh 43  
 exponentielle 74, 77, 240  
 Gamma 250  
 gaussienne 20, 29, 30, 33, 75, 77, 109, 250  
 laplacienne 141, 259  
 log-normale 240  
 $M$ -aire symétrique 19, 27, 73, 77, 92  
 uniforme 20, 29, 74, 259  
 uniforme (équiprobable) 18, 73  
 divergence 30, 68  
 à la normale *voir* : non-gaussianité  
 conditionnelle 103  
 développement de la — 102  
 localement symétrique 108  
 paramétrique 105  
 réduite par traitement 104  
 symétrique 33  
 dynamique 74, 256

**E**

écart entropique *voir* : divergence  
 effacement 41, 60, 255  
 efficacité spectrale 172, 205  
 entropie 23, 54  
 binaire (fonction d'—) 26  
 changement d'échelle 29, 122, 127  
 conditionnelle 52, 54  
 conjointe 51  
 de Boltzmann 26, 245  
 définition axiomatique 239  
 développement de l'— 91  
 différentielle 28, 60, 63  
 différentielle conditionnelle 60, 64  
 d'ordre  $n$  160, 181, 269, 270  
 d'une source 160, 181, 193, 209, 269  
 d'un mélange 69, 125, 241, 246  
 et changement de variable 28  
 et erreur quadratique moyenne  
 minimale 143  
 et information de Fisher 142  
 maximale 25, 33, 67, 71, 73, 79

moyenne 240  
 nulle 24, 54  
 relative *voir* : divergence  
 sous-additivité 55  
 équipartition asymptotique (AEP) 213  
 équivoque 52  
 erreur  
 d'estimation 110  
 probabilité d'— 38, 39, 73, 78, 168, 173, 188, 198  
 quadratique moyenne (e.q.m.) 79, 168, 200  
 quadratique moyenne minimale (e.q.m.m.) 116, 118, 123  
 espérance 23  
 conditionnelle 117  
 estimation  
 dans du bruit gaussien 123  
 efficace 112  
 linéaire 118, 237  
 non biaisée 110, 118  
 optimale 112, 117, 119  
 paramétrique 105, 110  
 Euler (constante d'—) 240  
 évanouissement *voir* : canal à —

**F G H**

Fano (inégalité de —) 78–80, 182, 198–200, 264  
 fiable *voir* : distorsion, quasi-fiable  
 filtre 154, 156, 157  
 Fisher *voir* : information, matrice  
 gaussien(ne) *voir* : canal, distribution  
 Gibbs (inégalité de —) 72  
 goulot d'étranglement 102  
 Hadamard (inégalité de —) 245  
 Hamming *voir* : distance, poids de —  
 Hartley *voir* : (capacité du) canal  
 uniforme  
 hessien 110

**I**

inconditionnellement sûr 272  
 indépendance *voir* : dépendance  
 information de Fisher *voir aussi* : matrice  
 changement d'échelle 122

développement de l'— 114  
 d'un bruit additif 140  
 et entropie 142  
 et erreur quadratique moyenne  
   minimale 123  
 et information mutuelle 137  
 inégalité de l'— 135, 137, 250, 253  
 non paramétrique 121, 137  
 paramétrique 105  
 perdue 115, 248  
 sensibilité au bruit additif 139  
 information (inégalité de l'—) 31, 68  
 information mutuelle 46, 51, 53, 64  
   auto-information 54, 66, 181  
   conditionnelle 92  
   conservée 100  
   développement de l'— 93  
   d'ordre  $n$  162  
   d'un mélange 127  
   et changement de variable 101  
   et erreur quadratique moyenne  
     minimale 144  
   et information de Fisher 137  
   maximale 71  
   minimale 70  
   perdue 99, 247  
   point-selle de l'— 247, 252  
   symétrique 97  
   transmise dans un canal 48, 58, 162

**J K L**

jacobien 28  
 Jensen (inégalité de —) 32, 67, 243  
 Karhunen-Loève (transformée de —) 30, 209  
 Kuhn-Tucker (conditions de —) 209  
 Kullback-Leibler *voir* : divergence  
 Lagrange (multiplicateur de —) 72, 208, 209  
 loi *voir aussi* : distribution  
   des grands nombres 211

**M**

Markov *voir* : chaîne  
 matrice de corrélation 76, 152  
   de covariance 21, 138

de transition 37  
 d'information de Fisher 109, 113, 121, 136  
 mélange (inégalité de —)  
   pour la variance entropique 131  
   pour la variance fishérienne 136  
   pour l'entropie 125, 251  
   pour l'erreur quadratique moyenne  
     minimale 134, 251  
   pour l'information de Fisher 133, 135, 251  
   pour l'information mutuelle 127, 251  
 mémoire *voir* : canal, source  
 Minkowski (inégalité de —) 245, 246  
 mot de canal 165  
   de code 166, 170, 172  
   de source 148, 165  
   d'information 172  
   typique 218, 219  
 moyenne *voir aussi* : Césaro, espérance  
 théorème de la — 62

**N O**

nat (*natural unit*) 25, 105, 137, 142  
 néguentropie 26  
 non-gaussianité 76, 247, 251  
 normal(e) *voir* : gaussien(ne)  
 norme euclidienne 150  
 observation 105, 116  
 OPTA (optimum de performance  
   théoriquement accessible) 179, 235, 273  
 orthogonalité (principe, condition d'—)  
   117, 119

**P**

paresse (éloge de la —) 235  
 poids de Hamming 81, 149, 169  
   maximal 189  
   minimal 189  
   moyen 76, 80, 168  
 probabilité *voir* : distribution de —, erreur  
   (probabilité d'—)  
 processus aléatoire 147  
   i.i.d. 149  
   stationnaire 148

puissance  
entropique *voir* : variance —  
moyenne 42, 75, 140, 144, 169

## Q R

quantification 61  
et divergence 104  
et information mutuelle 100  
uniforme 65, 273  
quasi-fiable 173  
rapport signal à bruit 42, 44, 140, 144, 171,  
173, 201, 206  
redondance 34, 240  
rendement *voir* : taux de codage de canal  
Riemann (somme de —) 63

## S

score développement du — 114  
identité de Blachman pour le — 250  
non paramétrique 121  
paramétrique 108  
Shannon  
*voir aussi* : théorème de —  
borne de — 259, 260  
borne sur l'entropie 76, 80, 202  
Claude Elwood 9  
formule de — *voir* : capacité du canal  
gaussien  
information de — *voir* : information  
mutuelle  
limites de — *voir* : capacité-coût,  
OPTA, taux-distorsion  
paradigme de — 9, 165  
unité d'information 25  
source 147  
*voir aussi* : distribution  
avec mémoire 150, 162, 194, 209, 220,  
269  
binaire 149, 151, 198, 235  
blanche 150  
gaussienne 150, 152, 200, 236  
*M*-aire 199  
markovienne 150–153  
sans mémoire 149, 161, 192, 197  
stationnaire 148, 269  
symétrique 207

Stam-Blachman *voir* : information de  
Fisher (inégalité de l'—)  
statistique suffisante 115  
Stein (identité de —) 250  
Stirling (formule de —) 241  
symbole 17, 147

## T U

taux de codage  
de canal 172  
de source 170  
source/canal 166, 167  
taux-distorsion (fonction —) 178–181,  
184, 186, 187, 191–195, 197  
théorème de Shannon  
pour l'échantillonnage 12, 205  
pour le codage de canal 183, 226  
pour le codage de canal sans contrainte  
de coût 184, 224  
pour le codage de source 181, 230  
pour le codage de source et de canal  
235  
pour le codage de source sans pertes  
181, 218, 220  
trace 136  
traitement 35  
conservant l'information 100  
déterministe 36, 37, 84, 90, 154  
discret 37  
global 85  
inverse 45, 101  
réciproque 44, 79, 80, 86  
traitement de données (théorème du —)  
pour la divergence 103, 247, 248  
pour l'erreur quadratique moyenne  
minimale 120, 248  
pour l'information de Fisher 113, 115,  
248  
pour l'information mutuelle 97, 98,  
128  
pour un traitement déterministe 99  
tri (algorithmes de —) 241  
typique  
dépendance — 215  
ensemble — 212, 214  
séquence — 212  
séquences conjointement —s 213

**V**

variable aléatoire *voir aussi* : distribution  
  binaire 18, 26  
  complexe 21  
  continue 20  
  déterministe 19, 22, 24  
  discrète 17  
  équivalente 24, 45  
  *M*-aire 17, 73  
  mixte 22  
  quantifiée 61  
  réduite 28, 122  
variance conditionnelle 118  
  développement des —s 118  
variance entropique 130, 132, 207, 208,  
  259, 260  
  changement d'échelle 130

  et variance 130  
  et variance fishérienne 143, 252  
  inégalité de la — 131, 246, 251–253,  
    260  
variance fishérienne 135, 136  
  changement d'échelle 135  
  et variance 135  
  et variance entropique 143, 252  
  inégalité de la — 135, 137  
Venn (diagramme de —) 55, 60, 95, 98, 99  
voie de retour *voir* : canal avec —  
vraisemblance (maximum de —) 222

**W X Y Z**

Wiener-Hopf (équations de —) 119  
Yule-Walker (équations de —) 119