

Probabilités et scrabble

Michel LAFOND
mlafond001@yahoo.fr

Résumé : Dénombrements, calculs de probabilités et usage des outils statistiques dans les tirages aléatoires de lettres au Scrabble.

Mots clés : Probabilités, Scrabble, statistique inférentielle, intervalles de confiance, simulation.

Rappelons que "Scrabbler" signifie faire un mot avec toutes les lettres du tirage (en principe de 7 lettres).

(Ce verbe figure dans l'ODS* 2012 mais pas dans le petit Larousse 2012).

* ODS = "Officiel du Scrabble" contient tous les mots autorisés avec de courtes définitions.

Il est tentant de se lancer dans le calcul de la probabilité de "scrabber" à partir d'un tirage aléatoire dans le sac plein. Mais attention aux pièges de l'équiprobabilité !

1 Dénombrement du nombre de tirages de 7 lettres possibles.

Le sac français contient 102 lettres dont : 15 E, 9 A, 8 I, 6 N, 6 O, 6 R, 6 S, 6 T, 6 U, 5 L, 3 D, 3 M, 2 B, 2 C, 2 F, 2 G, 2 H, 2 P, 2 V, 2 jokers (notés *) et 1 J, 1 K, 1 Q, 1 W, 1 X, 1 Y, 1 Z.

On fonctionne donc avec un alphabet de 27 lettres, le joker, considéré comme une lettre (appelé parfois lettre blanche), est la 27^{ème} lettre. On suppose que le tirage est effectué aléatoirement sans remise dans le sac (ou par informatique de plus en plus souvent maintenant) et que tout tirage est accepté.

En effet, dans la plupart des parties, on refuse les tirages qui, par exemple, ne contiennent pas au moins 2 voyelles et au moins 2 consonnes (en tous cas dans les premiers tours).

Combien y-a-t-il de tirages distincts de 7 lettres ? Voilà un bon exercice pour commencer.

Pour l'instant, les 15 E, les 9 A etc. sont indiscernables (comme c'est le cas dans la réalité).

On définit un tirage de type $(x_1, x_2, x_3, \dots, x_k)$ comme un ensemble composé d'un mot de x_1 lettres L_1 , d'un mot de x_2 lettres L_2 , --- Les lettres L_1, L_2 --- étant distinctes et les effectifs étant décroissants au sens large.

Ainsi, {AA, C, D, EEE} est un tirage de type (3, 2, 1, 1).

Dans ces conditions il y a 2 954 029 tirages distincts de 7 lettres.

Pour le démontrer, répartissons les tirages selon leurs types en utilisant le petit tableau ci-dessous :

Effectif de la lettre	Lettres concernées	Nombre de lettres concernées	cumul
15	E	1	1
9	A	1	2
8	I	1	3
6	N O R S T U	6	9 *
5	L	1	10
3	D M	2	12
2	B C F G H P V *	8	20
1	J K Q W X Y Z	7	27

* Ainsi, on lit ci-dessus qu'il y a 9 lettres dont l'effectif est au moins égal à 6.

Si C_n^p est le nombre de combinaisons de p éléments parmi n , voici le décompte :

type	commentaire	exemple	dénombrement
(7)	7 lettres identiques donc 7 E, 7 A ou 7 I	{SSSSSSS}	3
(6, 1)	6 lettres identiques + 1 lettre distincte	{EEEEEE, Z}	$9 \times (27 - 1) = 234$
(5, 2)	5 lettres identiques + 2 lettres identiques distinctes de la première	{PPPPP, SS}	$10 \times (20 - 1) = 190$
(5, 1, 1)	5 lettres identiques + 2 lettres distinctes entre elles et de la première	{UUUUU, C, A}	$10 \times C_{26}^2 = 3250$
(4, 3)	4 lettres identiques + 3 lettres identiques distinctes de la première	{TTTT, DDD}	$10 \times (12 - 1) = 110$
(4, 2, 1)	4 lettres identiques + 2 lettres identiques distinctes de la 1ère + 1 lettre	{AAA, OO, W}	$10 \times (20 - 1) \times (27 - 2) = 4\ 750$
(4, 1, 1, 1)	4 lettres identiques + 3 lettres distinctes entre elles et de la 1ère	{UUUU, *, E, G}	$10 \times C_{26}^3 = 26\ 000$
(3, 3, 1)	2 fois 3 lettres identiques + 1 lettre distincte	{RRR, SSS, E}	$C_{12}^2 \times (27 - 2) = 1\ 650$
(3, 2, 2)	3 lettres identiques + 2 fois 2 lettres identiques distinctes des autres	{RRR, **, VV}	$12 \times C_{19}^2 = 2\ 052$
(3, 2, 1, 1)	3 lettres identiques + 2 lettres identiques + 2 lettres distinctes	{RRR, EE, W, M}	$12 \times 19 \times C_{25}^2 = 68\ 400$
(3, 1, 1, 1, 1)	3 lettres identiques + 4 lettres distinctes entre elles et de la 1ère	{SSS, R, E, Z, A}	$12 \times C_{26}^4 = 179\ 400$
(2, 2, 2, 1)	3 fois deux lettres identiques + 1 lettre distincte	{MM, BB, EE, Z}	$C_{20}^3 \times (27 - 3) = 27\ 360$
(2, 2, 1, 1, 1)	2 fois deux lettres identiques + 3 lettres distinctes	{MM, BB, P, R, Z}	$C_{20}^2 \times C_{25}^3 = 437\ 000$
(2, 1, 1, 1, 1, 1)	1 fois deux lettres identiques + 5 lettres distinctes	{CC, B, A, P, R, Z}	$20 \times C_{26}^5 = 1\ 315\ 600$
(1, 1, 1, 1, 1, 1, 1)	7 lettres distinctes	{A, B, C, D, E, F, G}	$C_{27}^7 = 888\ 030$

Avec un total de 2 954 029 tirages possibles de 7 lettres.

Si on imposait au moins 1 voyelle et au moins 1 consonne, ce nombre serait de 2 424 847.

Si on imposait au moins 2 voyelles et au moins 2 consonnes, ce nombre serait de 1 489 775.

Mais attention ! Ces 2 954 029 tirages ne sont pas équiprobables.

Par exemple, il est évident que le tirage EELRSTT (qui permet de scrabbliser LETTRES) est bien plus fréquent que le tirage JKQWXYZ qui ne permet pas de scrabbliser grand-chose en dehors de la Pologne.

Un calcul simple $[C_{15}^2 \times 5 \times 6 \times 6 \times C_6^2]$ montre, compte tenu des effectifs des lettres, que le tirage EELRSTT est 283500 fois plus fréquent que JKQWXYZ !

Le calcul de la probabilité de scrabbliser avec un tirage aléatoire de 7 lettres du sac ne sera donc pas facile, et nous commencerons par examiner une situation plus simple.

2 Probabilité de scrabbliser avec un tirage de DEUX lettres.

Dans toute la suite, on appellera "TIRAGE" un mot dont les lettres sont classées par ordre alphabétique.

On dira qu'un tirage est BON s'il permet de scrabbliser.

Calculons la probabilité de scrabbliser au premier tirage, c'est-à-dire la probabilité qu'une anagramme du tirage soit l'un des 80 mots de deux lettres autorisés au Scrabble en 2013.

Rappelons que les lettres sont classées par ordre alphabétique, que "*" désigne un joker. "?" désignera une lettre quelconque parmi les 27.

Dans ces conditions, il y a 371 tirages (non équiprobables) de 2 lettres se décomposant ainsi :

27 tirages A? (de AA à A*), 26 tirages B? (de BB à B*), 25 tirages C?, 24 tirages D?, 23 tirages E?, 22 tirages F?, 21 tirages G?, 20 tirages H?, 19 tirages I?, 17 tirages J? (de JK à J* car il n'y a pas le tirage JJ), 16 tirages K?, (de Kl à K*), 16 tirages L? (de LL à L*), 15 tirages M?, 14 tirages N?, 13 tirages O?, 12 tirages P?, 10 tirages Q? (de QR à Q*), 10 tirages R?, 9 tirages S?, 8 tirages T?, 7 tirages U?, 6 tirages V?, 4 tirages W?, 3 tirages X?, 2 tirages Y?, le tirage Z* et le tirage **.

On vérifie facilement en examinant la liste des 80 mots de 2 lettres autorisés que parmi ces 371 tirages, il y en a exactement 88 qui scrabblent, ce sont :

AA, AB, AC, AD, AF, AH, AI, AK, AL, AM, AN, AR, AS, AT, AU, AV, AY, A*, BE, BI, BU, B*, CE, CI, CO, C*, DE, DO, DU, D*, EH, EJ, EL, EM, EN, ER, ES, ET, EU, EV, EX, E*, FI, F*, GO, G*, HI, HO, H*, IL, IM, IN, IP, IR, IS, IX, I*, J*, K*, LU, L*, MO, MU, M*, NO, NU, N*, OR, OS, OT, OU, O*, PU, P*, RU, R*, SU, SV, S*, TU, T*, UV, UW, U*, V*, W*, Y*, **.

Liste des 80 mots de 2 lettres autorisés au Scrabble (en 2013) :

AA AH AI AN AS AU AY BA BÊ BI BU ÇA CE CI DA DE DO DU EH
 EN ES ET EU EX FA FI GO HA HÉ HI HO IF IL IN JE KA LA LE LI
 LU MA ME MI MU NA NE NI NÔ NU OC OH OM ON OR OS OU PI
 PU RA RÉ RI RU SA SE SI SU TA TE TO TU UD UN US UT VA VÉ
 VS VU WU XI

Remarque : AA, BA, KA, BE, TO, VS, WU sont des mots bien connus des scrabbleurs.

Mais la probabilité de "scrabber" au premier tirage n'est pas égale à $\frac{88}{371} = 0,237 \dots$ car on n'a pas l'équiprobabilité. Pour rectifier le tir, la seule façon est de distinguer (par des indices par exemple) les 15 E, les 9 A, etc. et de pondérer les possibilités des bons tirages.

Ainsi, il y a désormais $C_{102}^2 = 5151$ tirages équiprobables parmi lesquels sont bons :

Les $C_9^2 = 36$ tirages AA (à savoir $A_1A_2, A_1A_3, \dots, A_8A_9$), les $9 \times 2 = 18$ tirages AB, ($A_1B_1, A_1B_2, \dots, A_9B_2$), ---, les 2 tirages Y* et le tirage ** des deux jokers.

Un décompte manuel est facile (un tableau sur une page suffit) et donne un cumul de 2 362 bons tirages équiprobables.

La probabilité de scrabber en tirant 2 lettres du sac plein est égale à $\frac{2\,362}{5\,151} = 0,458$.

J'ai calculé que la probabilité de scrabber avec un tirage de 3 lettres est environ 0,463. Il faut pour cela connaître la liste des 610 mots de 3 lettres autorisés en 2013 (donnée en annexe) et examiner un par un les 3 457 tirages possibles. Quelques heures suffisent. Le même travail serait difficile manuellement à partir de 4 lettres, à moins de travailler en équipe...

3 Probabilité de scrabber avec un tirage de SEPT lettres.

En tapant "scrabble" dans le moteur de recherche de Wikipédia, on trouve un long article dans lequel on peut lire :

Au total, dans l'édition 2012 (ODS 6), 386 264 mots sont admis, dont 80 de 2 lettres, 610 de 3 lettres, 2 509 de 4 lettres, 7 645 de 5 lettres, 17 318 de 6 lettres, 31 070 de 7 lettres et 46 329 de 8 lettres.

Pour calculer la probabilité de scrabber avec un tirage de 7 lettres, on pourrait procéder de la même manière que dans le paragraphe précédent pour un tirage de 2 (ou 3) lettres, mais il faudrait la liste des 31 070 mots de 7 lettres.

Je n'ai pas trouvé cette liste sur Internet, mais même si c'était le cas, encore faudrait-il pouvoir utiliser le fichier dans un programme. Il y a $C_{102}^7 = 18\,466\,953\,120$ tirages équiprobables de 7 lettres, et pour un ordinateur actuel, le calcul exact du nombre de bons tirages est parfaitement envisageable.

Ne disposant pas du fichier des 31 070 mots de 7 lettres, on ne peut pas calculer la fréquence exacte des bons tirages, mais on va l'évaluer en utilisant les gros outils de la statistique.

Considérons l'ensemble des 18 466 953 120 tirages équiprobables comme une population au sens statistique.

Un élément quelconque de cette population (un tirage) est bon (SUCCES) ou non. Un tirage sans remise de 7 lettres du sac plein constitue donc une épreuve de Bernoulli pour laquelle la probabilité du succès est le nombre p qu'on aimerait bien connaître.

Considérons maintenant un échantillon aléatoire de 300 tirages du sac plein (J'ai construit un tel échantillon par informatique pour éviter les biais du tirage manuel, à l'aide d'un programme donné en annexe). Pour chaque tirage de cet échantillon, cherchons sur un site (donné en annexe) s'il existe une anagramme autorisée au Scrabble.

(Lorsqu'il y a un joker c'est pénible car il faut essayer dans le pire des cas les 26 possibilités de remplacement, le logiciel n'admettant pas le joker comme lettre).

J'ai trouvé 71 bons tirages sur les 300 tirages de l'échantillon.

Ce qu'on appelle en statistique inférentielle la fréquence échantillon, c'est-à-dire la proportion de SUCCES dans l'échantillon est ici $f = \frac{71}{300} = 0,2366 \dots$ et la théorie nous dit que, puisque la taille n de l'échantillon est supérieure à 30 (ici $n = 300$), la fréquence théorique F des succès (c'est-à-dire la variable aléatoire égale à la fréquence d'un échantillon aléatoire) suit à peu près la loi normale $N\left(f; \sqrt{\frac{f(1-f)}{n-1}}\right)$ soit ici $N(0,237; 0,0246)$.

[Sa moyenne est $m = 0,237$ et son écart-type est $\sigma = 0,0246$].

La théorie nous dit enfin que l'intervalle de confiance de p à 95% est :

$$[m - 1,96 \sigma; m + 1,96 \sigma] = [0,189; 0,285].$$

Cela signifie qu'il y a 95% de chances que la probabilité p de scrabbliser en tirant 7 lettres (sans remise) du sac plein soit dans l'intervalle $[0,189; 0,285]$.

En d'autres termes, il y a **un peu moins d'une chance sur 4 de scrabbliser lors du premier tirage au Scrabble**, ce que l'on constate en jouant (en tout cas les très bons joueurs).

Si on élimine les tirages qui ont trop de consonnes ou trop de voyelles, cette probabilité augmente, mais je n'ai pas fait les calculs. Par ailleurs, tous les 4 ans, le vocabulaire autorisé au Scrabble est revu (en ajoutant certains des nouveaux mots qui apparaissent perpétuellement dans tous les pays francophones) et toutes les probabilités calculées dans cet article devront être revues à la hausse prochainement.

4. Annexe.

- De nombreux sites Internet prétendent trouver les anagrammes d'une liste de lettres donnée, mais la plupart utilisent une base de données très insuffisante. J'ai trouvé un seul site (incontournable pour les mots croisés) permettant (entre autres) de trouver absolument toutes les anagrammes d'une liste de lettres. C'est :

<http://www.mots-croises.ch/>

Dans la page d'accueil, taper "Dictionnaires" puis "Recherche d'anagrammes---"

Dans la case, entrez vos lettres, et vous aurez en quelques secondes, si elles existent, toutes les anagrammes, y compris les noms propres. Comme le vocabulaire utilisé est considérable (plus de 700.000 mots !), pour le problème qui nous intéresse, il faut que vous vérifiiez vous-même (avec le dictionnaire officiel ou avec un logiciel) que ces anagrammes sont bien autorisées au Scrabble. On trouve aussi dans le commerce pour environ 60 euros, une petite "calculatrice" dédiée au Scrabble qui donne pour chaque liste de lettres entrées les mots autorisés réalisables avec ces lettres.

- Ci-dessous un programme (MAPLE) permettant de simuler n tirages aléatoires sans remise de lm lettres chacun :

```
#maple16-jeux-scrabble-proba-01
```

```
sac := array(1..102) : a := array(1..102) :
```

```
listemots := NULL :
```

```
lm := 7 : # longueur des mots
```

```
sac := [E, E, E, E, E, E, E, E, E, E, E, E, E, E, E, E, A, A, A, A, A, A, A, A, A,
A, I, I, I, I, I, I, I, I, N, N, N, N, N, N, O, O, O, O, O, O, R, R, R, R, R,
R, S, S, S, S, S, S, T, T, T, T, T, T, U, U, U, U, U, U, L, L, L, L, L, D,
D, D, M, M, M, B, B, C, C, F, F, G, G, H, H, P, P, V, V, J, K, Q, W,
X, Y, Z, 'jok', 'jok'] :
```

```
n := 100 : #nb de tirages de lm lettres,
```

```
for i to n do imot := NULL : for j to 102 do a[j] := j : od:
```

```
for k from 1 to lm do kk := 103 - k :
```

```
h := floor(rand( ) · 1e-12 · kk + 1) : imot := imot, a[h] :
```

```
t := a[kk] : a[kk] := a[h] : a[h] := t :
```

```
od:
```

```
imot := [imot] : mot := NULL : for j to lm do mot := mot,
sac[imot[j]] : od:
```

```
listemots := listemots, [mot] :
```

```
od:
```

```
print(listemots) :
```

- *Liste des 610 mots de 3 lettres (autorisés en 2013)*

AAS	ACE	ADA	ADO	AGA	ÂGE	AGI	AÏD	AIE	AIL	AIR	AIS
AIT	ALE	ALU	ÂME	AMI	ANA	ÂNE	ANI	ANS	API	ARA	ARC
ARE	ARS	ART	ASA	ASE	AUX	AVE	AXA	AXE	AYS	BAC	BAH
BAI	BAL	BAN	BAR	BAS	BÂT	BAU	BÉA	BEC	BÉE	BEL	BEN
BER	BEY	BIC	BIO	BIP	BIS	BIT	BLÉ	BOA	BOB	BOF	BOG
BOL	BON	BOP	BOT	BOX	BOY	BRU	BUE	BUG	BUN	BUS	BUT
BYE	CAB	CAF	CAL	CAP	CAR	CAS	CEP	CES	CET	CHU	CIF
CIL	CIS	CLÉ	COB	COI	COL	COM	CON	COQ	COR	COU	COX
CRÉ	CRI	CRU	CUL	CUT	DAB	DAH	DAL	DAM	DAN	DAO	DAW
DEB	DÉO	DER	DES	DEY	DIA	DIN	DIS	DIT	DIX	DOC	DOL
DOM	DON	DOP	DOS	DOT	DRU	DRY	DUB	DUC	DUE	DUO	DUR
DUS	DUT	DZO	EAU	ÉCO	ÉCU	EGO	ÉLU	ÉMU	ÉON	ÉPI	ÈRE
ERG	ERS	EST	ÊTA	ÉTÉ	EUE	EUH	EUS	EUT	EUX	ÉWÉ	EXO
FAC	FAF	FAN	FAQ	FAR	FAT	FAX	FÉE	FER	FEU	FEZ	FIA
FIC	FIÉ	FIL	FIN	FIS	FIT	FLA	FOB	FOC	FOG	FOI	FOL
FON	FOR	FOU	FOX	FUI	FUN	FUR	FUS	FUT	GAG	GAI	GAL
GAN	GAP	GAY	GAZ	GEL	GÉO	GEX	GIN	GIS	GÎT	GLU	GOI
GON	GOS	GOY	GRÉ	GUÉ	GUI	GUR	GUS	GYM	HAI	HAN	HEM
HEP	HEU	HIA	HIC	HIE	HIP	HIT	HOP	HOT	HOU	HUA	HUB
HUE	HUI	HUM	HUN	IBN	IBO	ICI	IDE	IFS	ÎLE	ILS	ION
IPÉ	IRA	IRE	ISO	IVE	IXA	IXÉ	JAB	JAM	JAN	JAR	JAS
JET	JEU	JOB	JUS	KAN	KAS	KAT	KÉA	KEN	KET	KHI	KID
KIF	KIL	KIP	KIR	KIT	KOB	KOÏ	KOP	KOT	KRU	KSI	KWA
KYU	LAC	LAD	LAI	LAO	LAS	LED	LEI	LEK	LEM	LES	LET
LEU	LEV	LEZ	LIA	LIE	LIN	LIS	LIT	LOB	LOF	LOG	LOI
LOS	LOT	LUE	LUI	LUO	LUS	LUT	LUX	LYS	MAC	MAI	MAL
MAN	MAO	MAS	MAT	MAX	MEC	MÉL	MÉO	MER	MES	MET	MIE
MIL	MIN	MIR	MIS	MIT	MIX	MMM	MOA	MOB	MOI	MOL	MON
MOR	MOS	MOT	MOU	MOX	MUA	MUE	MUG	MUR	MUS	MUT	MYE
NAC	NAN	NAY	NÉE	NEF	NEM	NÉO	NES	NET	NEY	NEZ	NIA
NIB	NID	NIE	NIF	NIM	NIT	NOM	NON	NOS	NUA	NUE	NUI
NUL	NUS	OBA	OBI	ODE	OFF	OHÉ	OHM	OIE	OÏL	OKA	OLA
OLÉ	ONC	ONT	OPE	ÖRE	ORS	OSA	OSE	OST	OTA	OTE	OUD
OUF	OUH	OUI	OUT	OVE	OXO	OYE	PAF	PAL	PAN	PAP	PAR
PAS	PAT	PEC	PEP	PET	PEU	PFF	PHI	PHÔ	PIC	PIE	PIF
PIN	PIS	PIU	PLI	PLU	POP	POT	POU	PRÉ	PRO	PSI	PST
PSY	PUA	PUB	PUE	PUR	PUS	PUT	PUY	QAT	QIN	QUE	QUI
RAB	RAC	RAD	RAI	RAM	RAP	RAS	RAT	RAY	RAZ	RÉA	RÉE
REG	REM	REZ	RHÉ	RHÔ	RIA	RIE	RIF	RIO	RIS	RIT	RIZ
ROB	ROC	ROI	ROM	ROS	ROT	RUA	RUE	RUS	RUT	RUZ	RYE
SAC	SAÏ	SAL	SAR	SAS	SAX	SEC	SEL	SEN	SEP	SES	SET
SIC	SIL	SIR	SIS	SIX	SKA	SKI	SOC	SOI	SOL	SOM	SON
SOT	SOU	SPA	SPI	SUA	SUC	SUD	SUE	SUP	SUR	SUS	SUT
TAC	TAF	TAG	TAN	TAO	TAR	TAS	TAT	TAU	TEC	TEE	TEK
TEL	TEP	TER	TES	TÊT	TEX	THÉ	TIC	TIF	TIN	TIP	TIR
TOC	TOF	TOI	TOM	TON	TOP	TOS	TÔT	TRI	TUA	TUB	TUE
TUF	TUS	TUT	UDS	UNE	UNI	UNS	URE	USA	USE	UTE	VAL
VAN	VAR	VAS	VAU	VER	VÉS	VÊT	VIA	VIE	VIF	VIL	VIN

VIS	VIT	VOL	VOS	VUE	VUS	WAD	WAP	WAX	WEB	WOH	WOK
WON	WUS	YAK	YAM	YEN	YET	YIN	YOD	YUE	ZEC	ZÉE	ZEF
ZEK	ZEN	ZIG	ZIP	ZOB	ZOÉ	ZOO	ZOU	ZUP	ZUT		