

# Vive les arbres ! ou l'art de représenter graphiquement

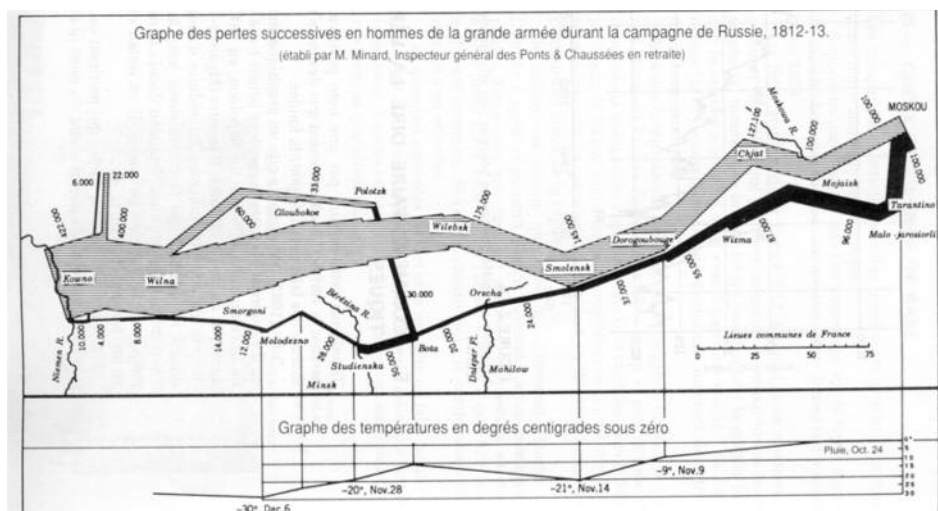
Rémy Coste

*Cet article a paru dans le n°112 des Chantiers de pédagogie mathématique, revue de la régionale Ile De France... ceux qui auraient encore le numéro constateront que le nom de l'auteur n'est pas le même ! En effet, Remy Coste avait signé sous le pseudo de Rachid Chibani de sorte que les lecteurs des Chantiers ne lisent pas toujours les mêmes auteurs... il n'était pas facile de recevoir des articles alors le comité de rédaction était sans cesse sollicité. Plot ne connaît pas trop ce problème et s'en réjouit. Si la première parution a eu lieu en mars 2002... le contenu de l'article est toujours d'actualité.*

C'est bien connu, la solution de certains problèmes en mathématiques réside dans l'art de les reformuler, ou dans celui d'en trouver une bonne représentation graphique. La théorie des graphes et son introduction dans le programme de spécialité en terminale ES, entièrement conçu autour de la résolution de problèmes, en est un parfait exemple. Dans les précédentes moutures des programmes, le recours aux arbres pour traiter un problème de probabilité avait déjà, à juste titre, été largement promu. Les arbres, outre une bonne façon de traiter certains problèmes, constituent également un outil pédagogique extrêmement efficace lorsqu'il s'agit de faire comprendre la notion

de probabilité conditionnelle, d'établir la formule  $p(A \times B) = p(A) \times p_A(B)$ , ainsi que sa sœur jumelle  $p(A \cap B) = p(B) \times p_B(A)$  obtenue en inversant l'ordre de construction de l'arbre, la formule des probabilités totales, ou encore la formule de la loi binomiale dans le cas où le nombre de répétitions est petit.

Certaines représentations graphiques sont restées célèbres. Dans l'excellent livre *Statistique – Économie-Gestion-Sciences-Médecine* de statistique de Thomas et Ronald Wonnacott (une véritable bible pour qui veut se former), on trouve le graphique ci-contre dont Tufte, en 1983, disait qu'il « pourrait bien être le meilleur jamais tracé ».



Il s'agit du graphique de la campagne de Russie de Napoléon, en 1812, établi cinquante ans après par un de ses officiers. Il montre comment une armée de près de 400 000 hommes fut réduite d'abord à 100 000 hommes durant les combats qui la conduisirent jusqu'à Moscou (bande grise). Trouvant la ville détruite, Napoléon décida la retraite qui se fit sous la neige et, à la frontière polonaise, il ne restait de la grande armée que 10 000 hommes (bande noire).

Ainsi, sur près de 40 hommes qui partirent en campagne ce matin de juin 1812, seulement un seul revint six mois plus tard, à moitié mort de froid.

En particulier, la tristement célèbre hécatombe provoquée par la traversée de la Bérézina par  $-20^{\circ}\text{C}$  apparaît de façon éclatante.

L'idée de représenter une grandeur par une branche dont la largeur est proportionnelle à son effectif est reprise dans le même ouvrage pour les arbres de probabilité illustrant la formule de Bayes. Et comme toutes les idées simples mais spectaculairement efficaces, on se demande pourquoi on ne l'a pas eue plus tôt.

Voici maintenant deux exemples d'utilisation de cette représentation.

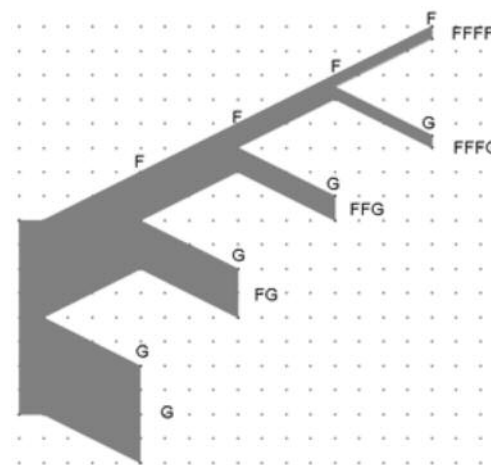
## Politique nataliste sexiste

Il s'agit de l'un des thèmes d'étude proposés dans l'ancien programme de seconde devenu un classique : une politique nataliste imaginaire impose à toute famille d'enfanter impérativement jusqu'à avoir un garçon, dans la limite maximale de 4 enfants, et de ne plus avoir d'enfant dès qu'un garçon est né. Il en découle que les seules fratries possibles sont constituées par 1 garçon (G), 1 fille et 1 garçon (FG), 2 filles et 1 garçon (FFG), 3 filles et 1 garçon (FFFG), ou bien 4 filles (FFFF). On

suppose qu'à chaque naissance, une fille a autant de chance de naître qu'un garçon. Le problème posé est le suivant : une telle politique favoriserait-elle une prédominance de filles ? de garçons ? ou bien conserverait-elle l'équilibre naturel ?

Pour un élève de seconde, il n'est pas intuitif de penser que les 5 configurations possibles ont des chances différentes d'apparaître. Quant à la question posée, l'expérience montre que la bonne réponse n'est spontanément intuitive pour personne, pas même un parterre de professeurs de mathématiques.

La réalisation de l'arbre ci-dessous dévoile en un clin d'œil les informations utiles pour donner la réponse et en établir une démonstration.



À chaque étape, l'arbre se scinde en 2 branches d'égales épaisseurs, traduisant ainsi l'équiprobabilité à chaque naissance. On voit parfaitement la constance de l'« équité » du partage des branches, mais aussi que l'épaisseur va ainsi s'amenuisant, ce qui illustre l'aspect de plus en plus rare de la configuration finalement obtenue. Tel un fleuve qui se subdivise régulièrement, le débit du plus lointain des bras est évidemment le plus faible.

La grille en arrière-plan du graphique permet de quantifier ce phénomène. Ainsi la fratrie FFFF a 8 fois moins de chance

d'apparaître que la configuration G, mais elle en a autant que la fratrie FFFG. Cela apparaît sans rien connaître des calculs de probabilités. Ainsi, si l'on prend comme distribution des fréquences celle qui est donnée par cet arbre (et c'est une belle occasion de réaliser une simulation pour observer la pertinence de ce choix), on peut évaluer le nombre moyen d'enfants par famille, le nombre moyen de filles et celui de garçons par famille (moyennes pondérées par les fréquences théoriques).

Nombre moyen d'enfants par famille :

$$\frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{4}{16} = \frac{30}{16} = 1,875.$$

Nombre moyen de garçons par famille :

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16} = 0,9375.$$

Nombre moyen de filles par famille :

$$\frac{1}{4} + \frac{2}{8} + \frac{3}{16} + \frac{4}{16} = \frac{15}{16} = 0,9375.$$

La conclusion s'impose : cette curieuse politique nataliste n'entraînerait pas un déséquilibre des sexes.

### Le jeu des trois portes

Il s'agit là aussi d'un grand classique. Lors d'un jeu de télévision célèbre aux Etats-Unis, un candidat fait face à 3 portes identiques. Derrière deux d'entre elles se

cache une chèvre, derrière la troisième se cache une automobile, constituant le cadeau convoité (et supposé préféré aux chèvres !). Le candidat désigne au hasard une des trois portes sans l'ouvrir (il ne sait donc pas ce qu'elle cache). Parmi les deux portes restantes, le meneur de jeu en élimine une derrière laquelle se trouve une chèvre (c'est le cas d'au moins une de ces deux portes). On demande alors au candidat de faire un choix définitif, entre la porte qu'il avait choisie en premier lieu et la troisième porte restante.

Il est assez simple de voir que s'offrent au candidat trois stratégies différentes : après la première étape, soit il choisit l'une des deux portes restantes au hasard, soit il en reste systématiquement à son premier choix, soit il change d'avis systématiquement. La question est de savoir quelle est la stratégie la plus favorable. Il est fortement intuitif de penser que dans la mesure où il ne reste que deux portes après la première phase, l'une étant gagnante et l'autre pas, toutes les stratégies se valent puisque l'on a 1 chance sur 2 de choisir finalement la bonne porte. Nous allons voir qu'il n'en est rien.

La petite histoire raconte qu'une polémique a pris une ampleur considérable



par journaux interposés, les partisans d'une stratégie s'opposant à ceux d'une autre. Une journaliste scientifique s'est même fait, paraît-il, prendre à partie de façon virulente, à la suite d'un article donnant une solution du problème, ce qui, en principe, aurait dû clore la discussion. Cela n'a pas été le cas, certains téléspectateurs continuant de privilégier leur intuition fautive (il y a une chance sur deux de gagner dans tous les cas) à un raisonnement logique, fut-il correct.

Pour tenter de faire mieux que la journaliste, utilisons les arbres « à branches à épaisseur ». Les trois graphiques ci-dessous représentent les différentes éventualités selon la stratégie utilisée. C1 et C2 sont les portes derrière lesquelles se trouve une chèvre, V celle derrière laquelle il y a la voiture (le candidat ne le sait pas). Si C1 a été d'abord désignée par le candidat, l'animateur éliminera la porte C2, le candidat aura alors à choisir entre C1 et V ; si c'est C2, l'animateur élimine C1, le candidat aura alors à choisir entre C2 et V. Si c'est V, l'animateur éliminera au choix C1 ou C2, la porte restante étant désignée indifféremment par C dans les graphiques.

Après avoir réalisé les 3 graphiques correspondants aux 3 stratégies possibles, il n'y a alors plus aucune difficulté pour conclure : la stratégie n° 3 est la meilleure, elle donne 2 chances sur 3 de gagner. La stratégie n° 2 est la moins bonne (1 chance sur 3 de gagner). Quant à la stratégie n° 1, elle donne 3 chances sur 6 de gagner, c'est-à-dire 1 chance sur 2. On peut même comprendre l'erreur que l'on fait en pensant intuitivement que le candidat aura toujours 1 chance sur 2 de gagner. Elle réside dans le fait que le choix final se fait entre deux portes dont on pense à tort qu'elles sont toutes les deux également prometteuses. Or, les arbres le montrent très clairement, dans tous les cas la porte qui avait été choisie en premier a 2 fois moins de chance de cacher la voiture qu'en a la troisième porte restante. Du coup il vient un raisonnement qui permet de résoudre le problème directement : lorsque l'on choisit dans un premier temps l'une des trois portes, il y a 1 chance sur 3 qu'elle cache la voiture. Donc, la dernière porte (celle qui reste après l'élimination effectuée par l'animateur) a forcément 2 chances sur 3 de cacher la voiture ; c'est donc elle qu'il faut choisir.

<p style="text-align: center;">Stratégie 1 Le choix final se fait au hasard, de façon équiprobable.</p>	<p style="text-align: center;">Stratégie 2 De façon déterminée, le 1<sup>er</sup> choix est conservé.</p>	<p style="text-align: center;">Stratégie 3 De façon déterminée, le 1<sup>er</sup> choix n'est pas conservé.</p>