par Jérôme Perez

Les naissances et les phases de la lune

Ce que disent les données

L'ère des données est arrivée. Encore faut-il savoir les analyser convenablement ! Selon une croyance populaire tenace, la lune a une influence sur les naissances. Est-ce vrai ? Que nous disent les données ? Réponse avec les mathématiques et la transformée de Fourier, impitoyable avec les séries périodiques.

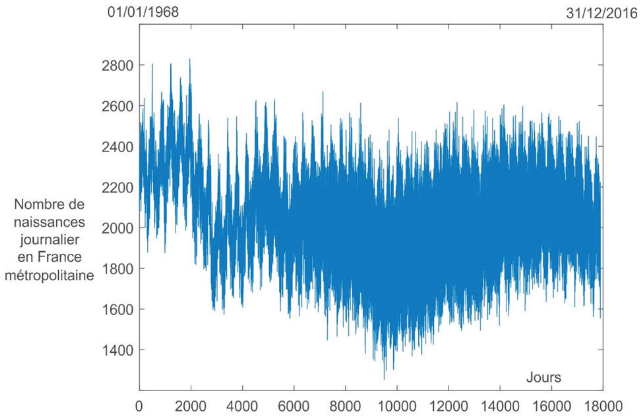
Depuis des temps immémoriaux, on attribue à la lune un « pouvoir » ou au moins une influence sur les naissances : une partie du corps médical en est convaincue (pratique à l'appui) et nombreuses sont les personnes qui pensent que la pleine lune favorise les naissances. L'Institut national de la statistique et des études économiques (Insee) fournit le nombre de naissances journalières en France métropolitaine entre le 1er janvier 1968 et le 31 décembre 2016 (**https://insee.fr**, onglet « Statistiques et études », puis rubrique « Les naissances en 2016 »). Cela représente plus de dix-sept mille cinq cents jours consécutifs et environ trente-neuf millions de naissances. Les données, significatives, se prêtent donc à toutes sortes d'analyses statistiques.

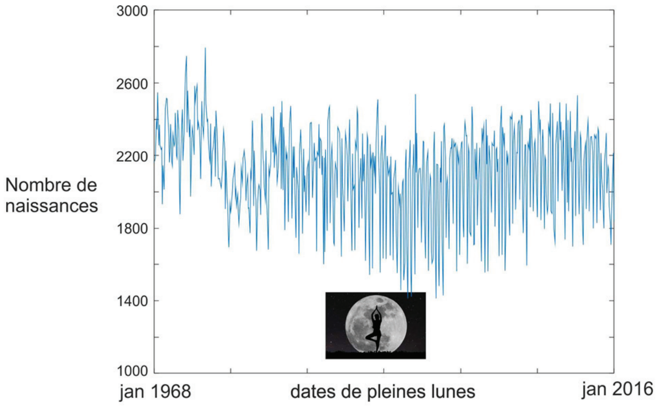
Les phases de la lune

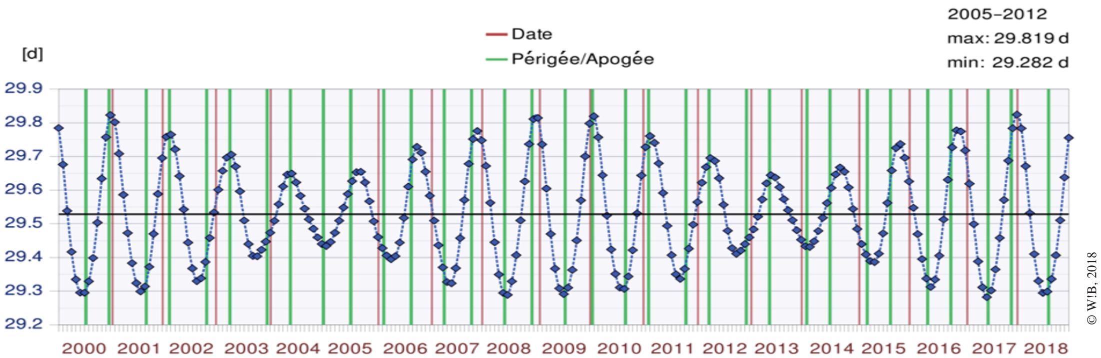
Commençons par tracer le nombre de naissances en fonction des jours (0 représente la date de début des enregistrements, à savoir le 1er janvier 1968). Le graphique, brut, n'est pas très parlant, même si l'on décèle, dans cette série, un certain caractère périodique, une moyenne de l'ordre de deux mille naissances par jour, et des fluctuations de l'ordre de plus ou moins cinq cents naissances autour de cette moyenne. Focalisons-nous alors sur les relations entre cette série et les « jours lunaires ». C'est en effet là que l'on prétend trouver quelque corrélation.

Les phases de la lune sont dues à l'éclairement de notre satellite par le soleil vu depuis la Terre. La lune tourne autour de notre planète en vingt-sept jours, sept heures et quarante-trois minutes en moyenne (cette période varie en fait légèrement : elle peut osciller de quelques dizaines de minutes à cause de diverses perturbations). Pendant cette période, nous nous sommes déplacés autour du soleil. Ainsi, depuis la Terre, il faut attendre « un peu plus » que vingt-sept jours pour revoir la lune dans la même phase ; c'est la *période synodique*, ou *lunaison*. Sa valeur moyenne est de vingt-neuf jours, douze heures et quarante-quatre minutes. La lunaison oscille également pour les mêmes raisons (influence du soleil et de la forme aplatie de la Terre, notamment).
Pour connaître exactement la phase de la lune chaque jour sur de longues périodes, il est nécessaire de prendre en compte toutes ces variations, sinon on peut se trouver confronté à de « petits » décalages locaux (mais pas globaux ou séculaires car la variation relative est nulle en moyenne). Les données de l'Institut de mécanique céleste et de calcul des éphémérides (IMCCE) prennent en compte toutes ces perturbations pour établir les phases lunaires.

Traçons les naissances uniquement les jours de pleine lune. On constate que le signal ressemble globalement à celui des données totales, ce qui est conforme à une distribution totalement aléatoire des naissances sur une année.







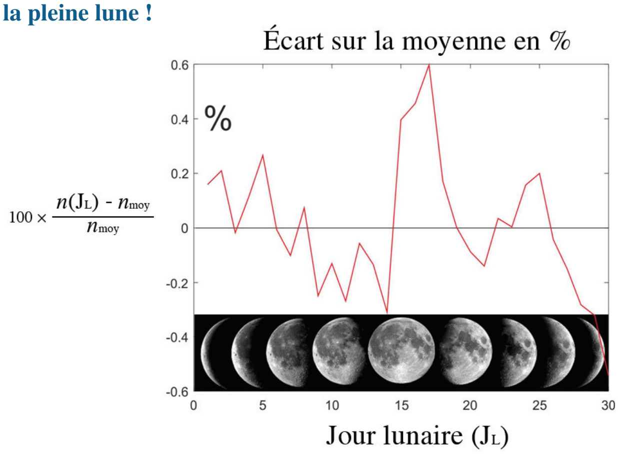
Variation de la durée de lunaison entre 2010 et 2018.

Il faut maintenant comparer les jours de pleine lune aux autres jours lunaires ! Une première analyse simple consiste à faire des moyennes. Grace aux données de l'IMCCE, chaque jour de l'intervalle d'étude T, qui s'étend du 01–01–1968 au 31–12–2016, peut se voir attribuer un « jour lunaire » J_L, identifié par un nombre entier compris entre 1 et 30, correspondant à la phase de la lune ce jour-là. Grace aux données de l'Insee, pour chaque jour lunaire de l'intervalle d'étude, on possède le nombre $m(i, J_L)$ de naissances, où i indique le nombre de jours lunaires du type J_L écoulés depuis le 01–01–1968 (i varie donc entre 1 et un peu plus de 580, au maximum). Soit $p(J_L)$ le nombre de fois où le jour lunaire J_L est survenu dans l'intervalle T. On peut calculer le nombre $N(J_L)$ de naissances total survenues un jour lunaire donné sur T :

$$\mathrm{N}(\mathrm{J_L}) = \sum_{i=1}^{p(\mathrm{J_L})} m(i, \mathrm{J_L}).$$

Le jour lunaire $i = 30$ possède un effectif réduit de moitié : pour avoir un trentième jour lunaire avec une période d'un peu plus de vingt-neuf jours et douze heures, il faut que le jour 29 se termine après le midi du jour calendaire…

Un pic est bien présent au moment de la pleine lune !



On peut donc aussi calculer $n(\mathrm{J_L}) = \dfrac{\mathrm{N}(\mathrm{J_L})}{p(\mathrm{J_L})}$, qui correspond à la moyenne des naissances chaque jour de la lunaison. Ce nombre moyen est de l'ordre de quelques milliers. On peut en évaluer les variations relatives. Il faut pour cela déjà calculer la moyenne $n_{\mathrm{moy}} = \dfrac{1}{30}\displaystyle\sum_{\mathrm{J_L}=1}^{30} n(\mathrm{J_L})$.

L'écart relatif entre $n(J_L)$ et sa moyenne n_{moy} en fonction du jour de lunaison est représenté ci-après en pourcentages. Si l'on ne lit pas en détail cette courbe, on pourrait penser que la lune a effectivement une influence sur les naissances, car on observe un pic aux alentours de la pleine lune !

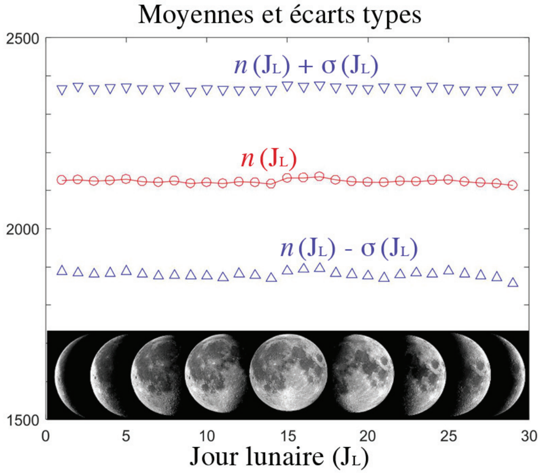
Fourier, le juge de paix

Le pic du schéma précédent est-il significatif ? Pour le savoir, il convient de prendre en compte les fluctuations des $n(J_L)$, qui sont par exemple décrites par l'écart type $\sigma(J_L)$ de ces quantités. Analyser sérieusement des données demande en effet du soin. On évalue donc $\sigma(J_L)$ par son estimateur standard (à savoir la racine carrée de la somme pondérée des carrés des écarts entre les données et leur moyenne) :

$$\sigma(\mathrm{J_L}) = \sqrt{\frac{1}{p(\mathrm{J_L})} \sum_{i=1}^{p(\mathrm{J_L})} \left(m(i, \mathrm{J_L}) - n(\mathrm{J_L})\right)^2}.$$

Cette quantité nous renseigne sur la façon dont les données sont « écartées » autour de la moyenne $\mu = \dfrac{1}{p(\mathrm{J_L})}\displaystyle\sum_{i=0}^{p(\mathrm{J_L})} n(\mathrm{J_L})$.

Si les données sont distribuées suivant une loi normale dans l'intervalle $[\sigma(J_L) - \mu, \sigma(J_L) + \mu]$ (ce qui serait le cas de données indépendantes et identiquement distribuées), on concentre environ 70 % des données. Pour avoir une idée de la signification des fluctuations observées sur la dernière courbe, il est donc opportun de tracer sur un même graphique $n(J_L)$, $n(J_L) - \sigma(J_L)$ et $n(J_L) + \sigma(J_L)$. On obtient des courbes plus représentatives du phénomène, car intégrant ces fluctuations. Les courbes sont « bien plus plates » et la lune ne semble pas avoir du tout d'influence significative sur les naissances. Pour en avoir définitivement le cœur net, il ne reste plus que l'analyse de Fourier, véritable juge de paix car impitoyable avec les séries périodiques. En effet, s'il existe une fréquence cachée dans un signal, sa transformée de Fourier la met immédiatement en évidence (voir encadré).

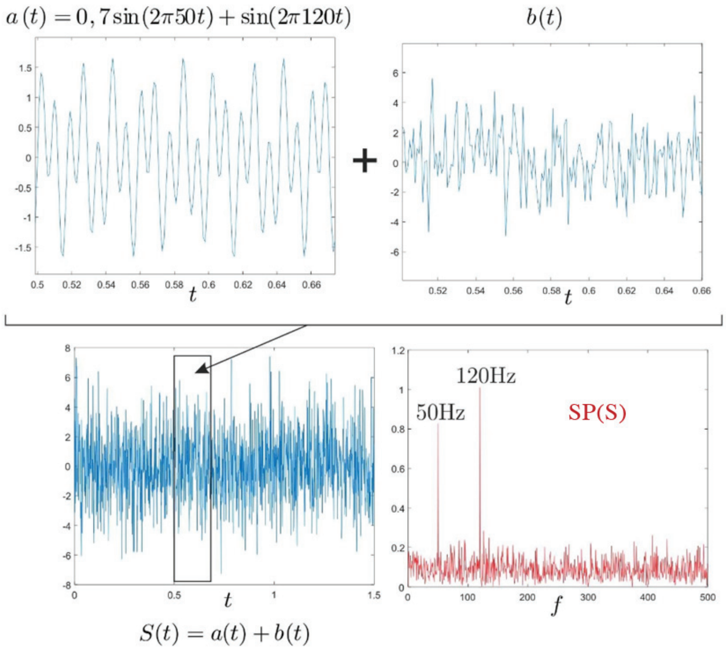


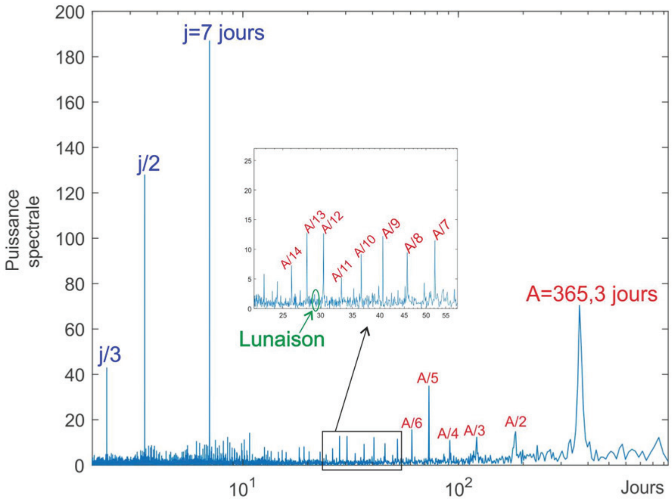
Un outil efficace

Prenons une fonction périodique a avec deux fréquences pures, $f_1 = 50$ Hz et $f_2 = 120$ Hz. On a donc $a(t) = 0{,}7 \sin(2\pi \times 50t) + \sin(2\pi \times 120t)$ avec le paramètre temporel t en secondes.

On ajoute à chaque point $a(t)$ de ce signal (échantillonné à 1 500 Hz) un nombre aléatoire $b(t)$ uniformément réparti entre -2 et 2. On ajoute ainsi un bruit b à la fonction a et on forme donc le signal total $S = a + b$.

Calculons le *spectre de puissance* SP(S) de ce signal bruité, à savoir la transformée de Fourier d'une quantité associée à S. Qu'observe-t-on ? Rien sur le signal S, mais son spectre de puissance traduit précisément la présence de la fonction a. On isole précisément les deux fréquences (il n'y a pas d'harmoniques car la fonction est composée de deux fréquences pures f_1 et f_2).

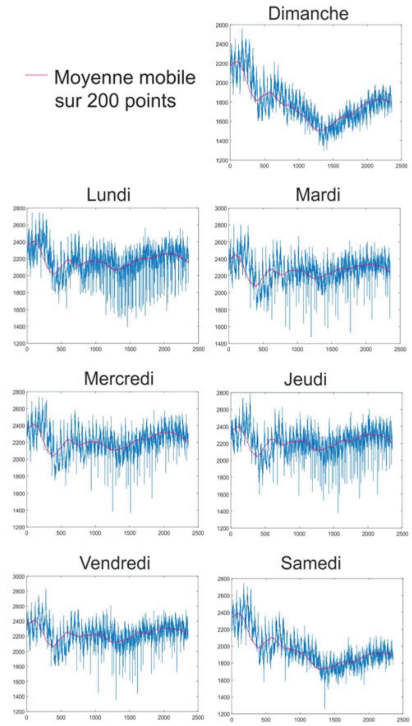




Transformée de Fourier du nombre de naissances par jour.

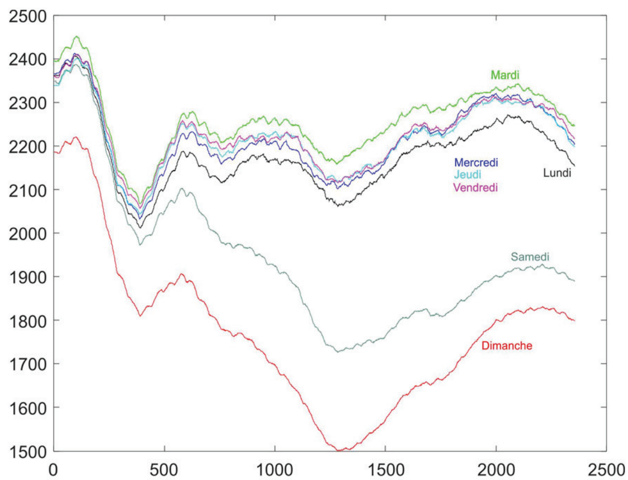
Représentons donc la transformée de Fourier du nombre de naissances journalier. En abscisse, sur le schéma, l'échelle temporelle est logarithmique. On remarque immédiatement deux périodes (et leurs harmoniques) : un signal de période A = 1 an (avec plus d'une dizaine d'harmoniques) et un signal de période j = 1 semaine (avec deux harmoniques). Par contre, on ne trouve rien du côté du cycle des phases lunaires, dont la période est d'environ 29,5 jours.

Pour comprendre l'origine de ces périodes, recensons le nombre de naissances chaque jour de la semaine. Les ordonnées de sept courbes ne sont pas identiques, afin d'obtenir des graphes tous lisibles. On constate que deux séries présentent une tendance globale à « diminuer » (le dimanche et le samedi) alors que les autres sont à peu près constantes tout au long de l'intervalle T.



Des réponses provenant du corps médical

Traçons sur un même graphique tous les jours de la semaine. Les dimanches sont systématiquement déficitaires, et les samedis le deviennent eux aussi six cents semaines après le 01–01–1968, soit vers le mois de juillet 1979.



On comprend alors que les médecins provoquent les accouchements de façon à éviter le dimanche, puis, à partir des années 1980, le samedi aussi… Peut-être cela correspond-il à la mise en service d'une nouvelle technique médicale ? En tous les cas, si chaque semaine il y a moins de naissances le week-end, cela explique la période j de sept jours. Le corps médical est astucieux et prévenant ! Il reste cependant à comprendre la période A d'un an. Pour cela, on recense grâce aux données de l'Insee le nombre moyen de naissances non pas chaque jour de la semaine, mais chaque jour de l'année. On trouve alors un graphique très intéressant.

On comprend tout ! Les jours fériés à date fixe, qui sont des sortes de « dimanches » distribués dans l'année, correspondent également, et très précisément, à des déficits de naissances. Le corps médical s'attribue ainsi des jours fériés en programmant les naissances. Une fois les données analysées, on est confronté à des phénomènes auxquels on ne s'attendait pas forcément, mais que l'on comprend en modifiant son point de vue.

On observe des pics de naissances début mai (malgré les ponts) et en septembre–octobre. En retirant la durée moyenne de la gestation d'une femme française (soit deux cent quatre-vingt-trois jours), on constate que les Français conçoivent plus de bébés pendant les vacances d'été et aux alentours de Noël : il serait intéressant d'étudier des données antérieures à 1936 et l'introduction des congés payés en France… Le pic de naissances de la fin d'année est sans doute en grande partie imputable au fait que l'on fait accoucher plus de femmes avant les vacances de fin d'année de façon préventive, car les hôpitaux sont alors privés d'une partie de leur personnel… La lune n'a donc rien à voir avec les naissances, malgré une croyance bien répandue ! Mais de toute façon, il est bien difficile de changer les croyances…

J.P.