

Annexe

Les *a priori* bayésiens de Laplace

En 1781, Laplace introduit la distribution *a priori* non constante des paramètres à estimer dans son principe de la probabilité des causes par les événements. Immédiatement, il se propose de montrer que celle-ci ne joue pas un grand rôle lorsque le nombre d'observations des événements considérés est grand (Laplace 1781, §. XXVIII, p. 469-470 ; M.-F. Bru et B. Bru 2018b, p. 379). Nous allons tenter d'explicitier ici sa démonstration.

Pré-requis

La quasi-totalité de ce qui suit est contenue dans l'article XVIII de son mémoire. Aussi, pour ne pas alourdir la notation, toute affirmation non référencée trouvera sa source dans les pages 469 et 470 de ladite publication (Laplace 1781). Par ailleurs, dans l'ensemble de cette démonstration, t désignera en nombre compris entre 0 et 1, mais ne valant ni 0, ni 1⁴⁸.

Laplace commence par poser deux *événements simples*⁴⁹ de probabilité inconnue respective x et $1 - x$. Il ne nomme pas ces événements, mais nous les appellerons E et F . Ainsi, $\mathbb{P}(E) = x$ et $\mathbb{P}(F) = 1 - x$. L'objectif est donc d'estimer x . Il pose ensuite deux fonctions, u et s , qu'il nomme les *facilités* des grandeurs x et $1 - x$. Il ne l'écrit pas comme cela, mais on peut définir ces fonction comme suit : $u(t) = \mathbb{P}(x = t)$ et $s(t) = \mathbb{P}(1 - x = t)$. Ce sont ces fonctions, supposées jusqu'ici constantes, que Laplace se propose pour la première fois de rendre quelconque. Afin de rendre la démonstration plus tangible, nous allons supposer que l'expérience a eu lieu $p + q$ fois, amenant p fois E et q fois F . Laplace pose également y « la probabilité de l'événement observé » (p. 469). Cet *événement observé* est à comprendre comme l'événement composé de p observations de E et q observations de F . Nous appellerons $E^p F^q$ cet événement⁵⁰. Aussi, même s'il n'en développe pas ici l'expression⁵¹, on peut raisonnablement écrire que $y(t) = \mathbb{P}(E^p F^q | x = t) = \binom{p+q}{p} t^p (1 - t)^q$. Dit autrement, y est la fonction de vraisemblance de l'événement $E^p F^q$. Tout l'enjeu de la probabilité des causes par les événements consiste alors à déterminer la probabilité de chacune

48. $t \in]0; 1[$.

49. Par *événement simple*, nous entendons un événement qui n'est pas constitué d'autres événements. Par exemple, lors d'un lancer de dé, si on estime qu'obtenir un nombre particulier est un événement simple, alors obtenir un nombre pair, impair ou encore premier, est un événement composé puisqu'il se constitue de plusieurs événements simples.

50. Il ne dit rien à ce sujet, mais il est implicite que les expériences sont indépendantes et, par conséquent, que l'ordre dans lequel les p E et les q F se produisent n'a aucune incidence sur le résultat. C'est aussi pour cela que nous nous permettons la notation $E^p F^q$.

51. Il donne plus haut (p. 423) la même expression sans coefficient binomial. Ce dernier importe cela dit assez peu dans notre situation étant donné qu'il n'est pas fonction de t ni de x .

des valeurs potentielles de x afin de trouver celles qui sont les plus probables étant donné les $p + q$ informations que nous avons. Selon son principe, Laplace écrit alors que cette probabilité, qui est notre *a posteriori* et que nous écrivons plus souvent $\mathbb{P}(x = t | E^p F^q)$, vaut⁵² $\frac{usy \, dx}{\int usy \, dx}$. Le dénominateur de cette fonction étant l'intégrale d'une fonction positive à une variable dont les bornes sont fixées (ici, 0 et 1), il est donc positif et constant. L'*a posteriori* et la vraisemblance pondérée de ses *a priori* sont donc proportionnels.

Cœur de la démonstration

Laplace cherche alors le maximum de l'*a posteriori*, ce qui revient à déterminer le maximum de usy étant donné la proportionnalité évoquée à l'instant. Pour ce faire, il dérive cette dernière expression, ce qui lui donne

$$\frac{d(usy)}{dt} = \frac{d(us)}{dt} \cdot y + \frac{dy}{dt} \cdot us \quad (2.1)$$

usy trouvant son maximum⁵³ à l'endroit où sa dérivée s'annule, son objectif devient alors de trouver où cette annulation se produit. Il opère alors une substitution déjà introduite quelques pages plus haut (§. XVIII, p. 423) :

$$y \, dt = \alpha z \, dy \quad (2.2)$$

Ici, α est défini comme $1/p$ ou comme μ/q (p. 423). Autrement dit, $p = 1/\alpha$ et $q = \mu/\alpha$. Tout l'intérêt de cette définition, et notamment de l'introduction de ce μ , c'est d'avoir un dénominateur commun à p et q , et une manière commode de les comparer : p est alors μ fois plus petit que q . Il règne néanmoins sur α et μ des contraintes assez fortes puisque p et q restent des nombres entiers. $z(t)$ quant à lui vaut exactement $\frac{t(1-t)}{1-(1+\mu)t}$ (p. 424)⁵⁴.

Une fois la substitution évoquée effectuée et le membre de gauche passé à droite et remplacé par 0, on obtient

$$\frac{d(us)}{dt}(t) \cdot y(t) + \frac{y(t)}{\alpha \cdot z(t)} \cdot u(t) \cdot s(t) = 0 \quad (2.3)$$

En divisant l'équation 2.3 par $u(t)$, $s(t)$, $y(t)$ et en la multipliant par α , on obtient

$$\alpha \frac{\frac{d(us)}{dt}(t)}{u(t) \cdot s(t)} + \frac{1}{z(t)} = 0 \quad (2.4)$$

Laplace estime ensuite que l'on peut négliger la fonction que multiplie α puisque ce dernier est, selon lui, un « coefficient très petit ». Il ne le détaille pas ici mais on peut constater que considérer α petit revient, par définition, à prendre p et q très grands⁵⁵. Or, c'est exactement

52. Notons qu'il précise systématiquement les bornes de ses intégrales dans le texte qui entoure les équations, mais ne les écrit pas sur ses $\int$. Aussi, on notera le caractère quelque peu impropre de la notation dx étant donné que x est notre paramètre inconnu.

53. Laplace n'évoque pas les situations où il y a plusieurs extrema.

54. Pour retrouver cette expression, on peut vérifier que $\frac{dy}{dt}(t) = \frac{1}{\alpha} \frac{1-(1+\mu)t}{t(1-t)} y(t)$. Il suffit ensuite de recourir à quelques manipulations algébriques dont une impropre revenant à multiplier chaque membre de l'équation par dt comme s'il s'agissait d'un nombre.

55. Pourvu que μ ne soit pas négligeable devant α en ce qui concerne q .

le cadre dans lequel il se place⁵⁶. Cette négligence faite, l'équation 2.4 devient alors à

$$\frac{1}{z(t)} = 0 \quad (2.5)$$

Laplace termine sa démonstration sur cette ultime équation, mais nous allons détailler un peu plus ce que cela signifie pour vraiment expliciter son résultat.

Si l'on remplace la fonction z par son expression, l'équation 2.5 devient

$$\frac{1 + (1 + \mu)t}{t(1 - t)} = 0 \quad (2.6)$$

On sait que pour qu'une fraction s'annule, son numérateur doit s'annuler sans que son dénominateur ne s'annule. Or ce dernier trouve ses racines en 0 et en 1, et nous avons justement exclus ces points de l'ensemble de définition de la variable t . Par conséquent, l'équation 2.6 revient à l'équation $1 + (1 + \mu)t = 0$ dont l'unique solution est $t = \frac{1}{1+\mu}$. Pour comprendre la signification de cette solution, divisons le numérateur et le dénominateur de cette fraction par α . On obtient alors, après simplification et mobilisation des relation entre p , q et α , que

$$t = \frac{p}{p + q} \quad (2.7)$$

Conclusion

Lorsque p et q sont grands, la valeur qui rend la fonction usy maximale est sensiblement celle qui rend y maximale, à savoir, la moyenne arithmétique des observations.

56. Il ne fait aucune supposition particulière sur les fonctions u et s , mais M.-F. et B. Bru se permettent de supposer que us et $d(us)/dt$ sont bornées pour donner plus de rigueur à cette démonstration (2018b, p. 379).