Signification, tests d'hypothèses et approche bayésienne

Hans Ivers, Ph.D.

École de psychologie, Université Laval

6 février 2020

Signification, tests d'hypothèses et approche bayésienne

Hans Ivers, Ph.D.

Plan de la présentation

Exemple : jeu de données

On réalise une étude comprenant deux conditions : un traitement expérimental (EXP) et une condition contrôle (CTL). Un total de $n = 10$ sujets par condition est recruté. On désire comparer la moyenne d'un indicateur Y pour chaque condition suite à l'étude (post-traitement).

Groupe	Moyenne	Écart-type	Étendue
Exp (n=10)	7.60	2.95	3-13
Ctl (n=10)	5.20	3.77	0-13

Exemple : test d'hypothèse

On réalise un test statistique (par exemple, un test t de Student) pour comparer les deux moyennes sur la variable dépendante Y.

Question

Quelle hypothèse nulle sera utilisée par le test d'hypothèse?

Exemple : test d'hypothèse

On réalise un test statistique (par exemple, un test t de Student) pour comparer les deux moyennes sur la variable dépendante Y.

Question

Quelle hypothèse nulle sera utilisée par le test d'hypothèse?

Réponse : l'effet du traitement est nul

$H_0 : \mu_1 = \mu_2 \implies \mu_1 - \mu_2 = 0$

Pour simplifier, on dira que la différence $\theta = 0$.

Petits rappels

- une *statistique* est une fonction des *observations*. Les statistiques les plus courantes sont la moyenne $\bar{x}$, la variance s^2 et les statistiques échantillonnales t de Student, F de Fisher (ANOVA), khi-carré χ^2 et la corrélation r de Pearson.
- un *paramètre* est une quantité (inconnue dans la population) qu'on désire estimer. Les paramètres les plus courants sont la moyenne "mu" μ, la variance "sigma carré" σ^2, la corrélation "rho" ρ et le coefficient de régression "beta" β. Pour simplifier la présentation, tout paramètre est appelé "theta" θ.

Exemple : test d'hypothèse

Une méta-analyse basée sur une trentaine d'études antérieures nous indique que la différence moyenne entre ces deux conditions est de 3 unités (variation entre 0 et 6 unités).

Question

Est-ce que ces connaissances vont modifier l'hypothèse nulle?

Exemple : test d'hypothèse

Une méta-analyse basée sur une trentaine d'études antérieures nous indique que la différence moyenne entre ces deux conditions est de 3 unités (variation entre 0 et 6 unités).

Question

Est-ce que ces connaissances vont modifier l'hypothèse nulle?

Réponse : non, les connaissances antérieures n'entrent pas dans le calcul du test d'hypothèse.

Exemple : valeur p

Le résultat du test t de Student se lit comme suit : $t(18) = 1.59, p = .13$.

Question

Comment interpréter la valeur p?

Exemple : valeur p

Le résultat du test t de Student se lit comme suit : $t(18) = 1.59, p = .13$.

Question

Comment interpréter la valeur p?

Réponse : sous $H_0(\theta = 0)$, la probabilité d'observer une statistique t à 18 degrés de liberté, égale ou supérieure à la valeur absolue de la statistique observée $\hat{t}$, est de $p = .13$. Interprétation : si H_0 est vraie, nous avons 13% de chances d'observer cette différence $(7.60 - 5.20 = 2.40)$ ou une différence plus extrême entre les deux moyennes de ce jeu de données.

Exemple : intervalle de confiance

Suite à ce test, on produit une estimation de la différence et de son erreur standard, $\hat{\theta} = 2.40$, $E.S. = 1.51$, ce qui donne un intervalle de confiance à 95% : $[-0.78, 5.58]$.

Question

Comment interpréter cet intervalle?

Exemple : intervalle de confiance

Suite à ce test, on produit une estimation de la différence et de son erreur standard, $\hat{\theta} = 2.40$, $E.S. = 1.51$, ce qui donne un intervalle de confiance à 95% : $[-0.78, 5.58]$.

Question

Comment interpréter cet intervalle?

Réponse :

$$Pr(IC_{inf} \leq \theta \leq IC_{sup}) = 1 - \alpha$$
$$Pr(-0.78 \leq \theta \leq 5.58) = 1 - 0.05 = 0.95$$

Interprétation : si on répète cette expérience et on recalcule cet intervalle une infinité de fois, $1 - \alpha$ % des intervalles vont contenir le paramètre de la population θ.

Implications de ces calculs

Quatre implications selon la théorie statistique classique :

- Il n'y a que deux conclusions : absence OU présence d'un effet, sans nuance...
- Le test statistique ne peut pas utiliser l'information antérieure disponible sur le paramètre;
- La valeur p d'une statistique ne nous renseigne pas sur la probabilité de $H_0(\theta = 0)$ ou $H_1(\theta > 0)$, mais plutôt sur la probabilité d'observer le jeu de données *SI* H_0 *est vraie*.
- 19 fois sur 20 (95%), le paramètre recherché est dans l'intervalle de confiance... mais l'intervalle calculé suite à notre test va contenir (probabilité $=$ 100%) OU non (probabilité $=$ 0%) le paramètre recherché.

Statistique classique

L'approche inférentielle dominante est basée sur les travaux de R. Fisher, inventeur du concept de vraisemblance (1922), et J. Neyman et E. Pearson, co-inventeurs de l'approche par test d'hypothèses (1933).

- θ est une constante inconnue (il existe une seule vraie valeur du paramètre dans la population)
- les observations sont aléatoires
- l'état du monde est divisé en deux possibilités mutuellement exclusives : hypothèse nulle ou alternative
- un seuil arbitraire ($\alpha = 5\%$) pour conclure au rejet de H_0 et déclarer un résultat "statistiquement significatif".

Dominance de l'approche classique

Pourquoi cette approche est toujours dominante, malgré des critiques en croissance (Task Force de l'APA, statement sur p-values de l'ASA)?

- la valeur p est un bon premier test pour distinguer les variations aléatoires de possibles effets réels;
- facilité de calcul et d'usage (implantation dans tous les logiciels);
- l'enseignement de la statistique est basé essentiellement sur ce paradigme;
- réforme difficile (alternatives plus complexes, majorité des usagers de statistiques ne sont pas statisticiens);

Tentatives de pallier aux limites

Différentes tentatives sont faites depuis 30 ans pour pallier aux limites de cette approche :

- mettre davantage l'accent sur la grandeur de l'effet et/ou les intervalles de confiance, au lieu de seulement rapporter la valeur p (Publication Manual de l'APA);
- utiliser un seuil de signification plus bas (p.ex., $p < .005$) (revues médicales);
- bannir les tests de signification (certains périodiques en psychologie);

Objectifs d'une approche idéale d'inférence

Dans une approche idéale d'inférence, on devrait (1) obtenir un maximum d'information sur le paramètre et non sur les données (nous les avons déjà en notre possession), et (2) avoir le choix d'intégrer de l'information antérieure, pour éviter de repartir à zéro pour chaque test, surtout dans les cas où un nombre limité d'observations est disponible.

Approche bayésienne

Caractéristiques de cette approche inférentielle :

- le paramètre inconnu θ n'est pas une constante mais une quantité aléatoire. On peut donc obtenir sa loi (sa distribution). Ce sont les observations qui sont fixes;
- on peut incorporer (ou non) l'information connue *a priori* sur le paramètre;
- l'inférence est réalisée par la description de la loi de θ (moyenne, intervalle de confiance, probabilité de prendre une certaine valeur, etc.), selon les données observées;
- peu d'intérêt pour un test d'hypothèse (quoiqu'il existe des outils pour le faire si besoin)

Théorème de Bayes

L'approche *bayésienne* tire son nom du révérend et mathématicien Thomas Bayes (1701-1761) qui s'intéressait aux "probabilités inversées".

Soit A et B deux événements. Si je connais la probabilité de A, $P(A)$, et la probabilité de B, sachant que A est survenu, $P(B|A)$, que puis-je déduire sur la probabilité inverse, soit $P(A|B)$?

Théorème de Bayes

Question

Qu'est-ce qu'une probabilité conditionnelle? p.ex. $P(B|A)$

En assumant que les événements A et B ne sont pas indépendants, c'est la probabilité de l'événement B si l'événement A est déjà survenu.

	B (oui)	B (non)
A (oui)	1	2
A (non)	3	4

Si A est survenu (n=3), B = oui est 1/3.
En langage probabiliste :
$P(B|A) = P(A \cap B)/P(A) = (1/10)/(3/10) = 1/3$

Théorème de Bayes

Formule du théorème de Bayes pour lier les deux probabilités inversées :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

Exemple d'usage du théorème de Bayes

On réalise un test par échographie pour savoir si un enfant à naître sera de sexe féminin. On sait que ce test est positif pour détecter une fille dans 98% des cas où l'enfant est une fille, mais il peut donner des faux positifs (20%) pour les garçons. De plus, on sait qu'il y a en moyenne 48 filles pour 52 garçons dans les naissances.

Question

Quelle est la probabilité d'avoir une fille si mon test est positif, $P(F|T^+)$?

Exemple d'usage du théorème de Bayes

Selon l'équation (1), avec A = avoir une fille et B = avoir un test positif :

$$
\begin{aligned}
P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\
P(F|T^+) &= \frac{P(T^+|F)P(F)}{P(T^+)} \\
&= \frac{0.98 \times 0.48}{0.98 \times 0.48 + 0.20 \times 0.52} \\
&= 0.8189
\end{aligned}
$$

Exemple d'usage du théorème de Bayes

On peut voir ce calcul comme une *mise à jour* de nos connaissances sur la probabilité d'avoir une fille.

$$P(F|T^+) = \left[\frac{P(T^+|F)}{P(T^+)}\right] \times P(F)$$

$$0.8189 = \left[\frac{0.98}{0.5744}\right] \times 0.48$$

- avant le test, on sait qu'on a 48% de chances d'avoir une fille (probabilité *a priori*)
- après le test positif (donc après avoir récolté des données), cette probabilité est passée à 82% de chances (probabilité *a posteriori*).

Concepts bayésiens

On peut tracer un parallèle entre cet exemple et une approche plus générale d'inférence :

- le test $=$ les données Y récoltées dans une étude
- le sexe $=$ le paramètre inconnu θ qu'on désire estimer

Donc, on peut définir l'approche bayésienne comme un **outil qui permet de calculer la loi du paramètre inconnu selon les données observées**. On qualifie cette loi de "conditionnelle aux données" ou loi "a posteriori".

Concepts bayésiens

Théorème de Bayes pour des événements simples:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Théorème de Bayes pour des distributions (lois), avec $A = \theta$ et B = observations Y:

$$loi(\theta|Y) = \frac{loi(Y|\theta)loi(\theta)}{loi(Y)}$$

Le calcul de la loi "a posteriori" est fonction de trois quantités : la vraisemblance des données, la loi a priori du paramètre et la loi des données (pas importante car sans lien avec θ).

Concepts bayésiens

La *loi*$(Y|\theta)$, appelée "vraisemblance des données", résume tout ce que l'échantillon nous apprend sur θ.

En inférence classique (les tests usuels), la méthode du *maximum de vraisemblance* utilise cette fonction pour déterminer la valeur de θ qui maximise cette fonction et permet donc de proposer un estimé $\hat{\theta}$ qui est le plus "vraisemblable" selon les observations récoltées.

Concepts bayésiens

La loi du paramètre lui-même, *loi*(θ), aussi appelée "loi a priori", se divise en deux catégories :

- une catégorie dite *informative* : on connait déjà la moyenne, ou l'étendue du paramètre, grâce à des études antérieures ou des jugements d'expert. Souvent, on utilise une loi symétrique de forme normale, centrée sur la valeur moyenne attendue.
- une catégorie dite *non-informative* : on ne sait rien à l'avance donc toute valeur est possible. Les lois les plus utilisées sont (1) la loi uniforme ou (2) la loi normale avec une variance très large (p.ex., $\sigma_0^2 = 10^6$).

Loi a posteriori de θ

Selon une approche d'inférence classique (ici, notre test t), qu'avons-nous appris sur θ?

- $\hat{\theta} = 2.40$, donc la valeur la plus vraisemblable de θ est 2.40
- il y a 95% de chances que la vraie valeur de la différence θ soit dans l'intervalle [-0.78, 5.58]

En stats bayésienne, nous avons accès à la distribution complète du paramètre (sa "loi a posteriori").

Loi a posteriori de θ

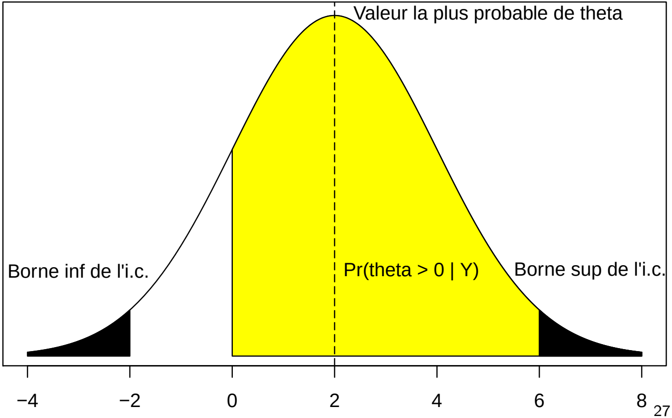
Que faire avec la "distribution d'un paramètre"?

A partir d'une distribution *a posteriori* du paramètre d'intérêt (ici, $\theta = \mu_1 - \mu_2$), on calcule généralement trois quantités :

- la valeur la plus probable de θ (la moyenne et/ou la médiane);
- intervalle de confiance à 95% (renommé intervalle de *crédibilité*) qui inclut 95% des valeurs possibles de θ;
- probabilité spécifique, par exemple que $\theta > 0$, ce qui donne une estimation de la probabilité de l'hypothèse alternative selon les données observées, $Pr(H_1|Y)$ que nous étions incapables d'obtenir en statistique classique

Loi a posteriori de θ

Illustration des quantités extraites de la loi a posteriori du paramètre :



Exemple d'analyse bayésienne

Analysons notre jeu de $n = 20$ observations divisées en deux conditions selon une approche bayésienne simple, en deux temps :

a) Approche avec loi a priori non-informative ($n = 20$)

vraisemblance des données = loi normale

loi a priori de la différence $\mu_1 - \mu_2 = \theta =$ loi normale avec moyenne 0 et variance $= 10^6$.

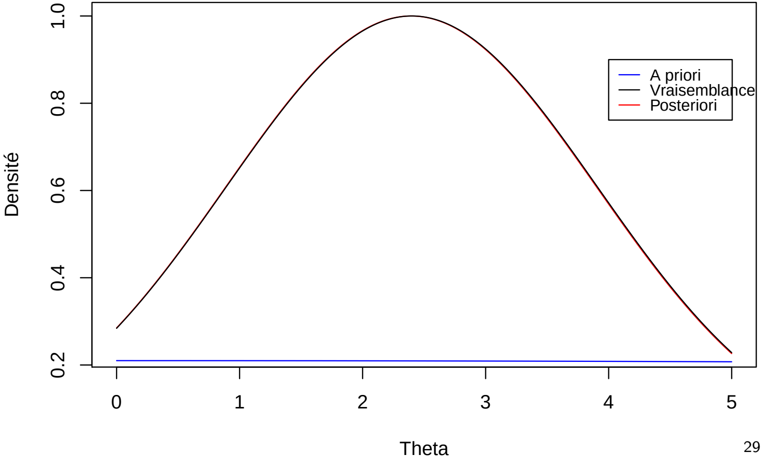
b) Approche avec loi a priori informative ($n = 20$)

vraisemblance des données = loi normale

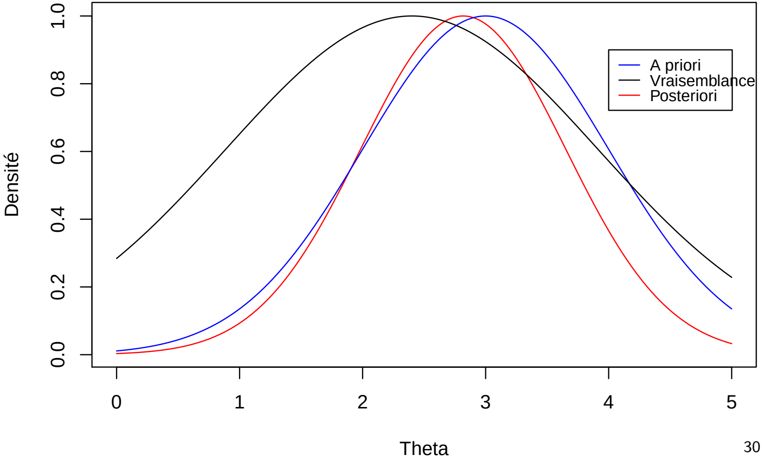
loi a priori de $\theta =$ loi normale de moyenne 3 et d'écart-type 1 (selon la méta-analyse).

Exemple d'analyse bayésienne

a) Approche avec loi a priori non-informative ($n = 20$)



b) Approche avec loi a priori informative ($n = 20$)



Exemple d'analyse bayésienne

Implications de ces résultats pour notre jeu de $n = 20$ observations :

a) Approche avec loi a priori non-informative

Lorsque nous n'avons aucune information a priori sur le paramètre, la loi a posteriori est dominée par la vraisemblance. Dans ce cas, les résultats de l'analyse bayésienne sont identiques à ceux avec l'approche statistique classique.

b) Approche avec loi a priori informative

La loi a posteriori du paramètre est une *moyenne pondérée* de la loi a priori (moyenne et variance) et de la vraisemblance (maximum et information).

Exemple d'analyse bayésienne

Ces résultats sont-ils comparables à ceux obtenus avec l'approche classique?

Approche	$\hat{\theta}$	I.C. 95%	Prob
Classique	2.400	-0.779, 5.579	p-valeur = .130
Bayes non-inf	2.400	-0.779, 5.578	$Pr(\theta > 0 \mid Y) = .943$
Bayes inf.	2.814	1.129, 4.360	$Pr(\theta > 0 \mid Y) = .999$

Note. $\hat{\theta}$ = différence estimée entre les deux conditions.

Exemple d'analyse bayésienne

L'analyse de la distribution a posteriori du paramètre nous permet de répondre facilement à des questions encore plus complexes (p.ex., avec la loi a priori informative):

- Quelle est la probabilité que la vraie différence entre les deux conditions θ soit entre 2 et 4 unités?

Exemple d'analyse bayésienne

L'analyse de la distribution a posteriori du paramètre nous permet de répondre facilement à des questions encore plus complexes (p.ex., avec la loi a priori informative):

- Quelle est la probabilité que la vraie différence entre les deux conditions θ soit entre 2 et 4 unités?
  Réponse : $Pr(2 \leq \theta \leq 4|Y) = 0.758$

Exemple d'analyse bayésienne

L'analyse de la distribution a posteriori du paramètre nous permet de répondre facilement à des questions encore plus complexes (p.ex., avec la loi a priori informative):

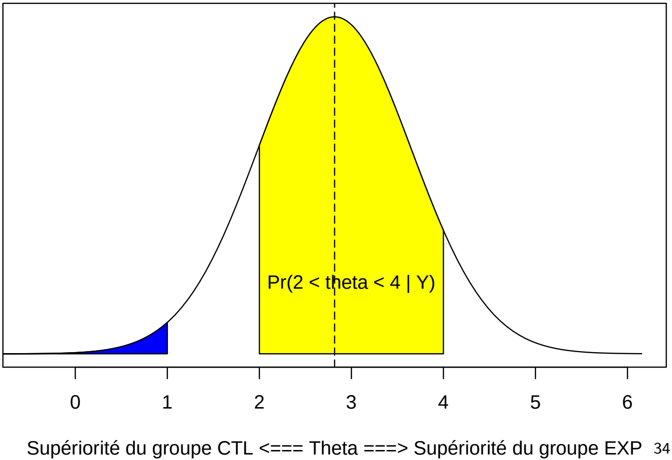
- Quelle est la probabilité que la vraie différence entre les deux conditions θ soit entre 2 et 4 unités?
  Réponse : $Pr(2 \leq \theta \leq 4|Y) = 0.758$
- Quelle est la probabilité que la vraie différence soit à l'intérieur d'une *marge clinique* de ± 1 unité? (devis de non-infériorité)

Exemple d'analyse bayésienne

L'analyse de la distribution a posteriori du paramètre nous permet de répondre facilement à des questions encore plus complexes (p.ex., avec la loi a priori informative):

- Quelle est la probabilité que la vraie différence entre les deux conditions θ soit entre 2 et 4 unités?
  Réponse : $Pr(2 \leq \theta \leq 4|Y) = 0.758$
- Quelle est la probabilité que la vraie différence soit à l'intérieur d'une *marge clinique* de ± 1 unité? (devis de non-infériorité)
  Réponse : $Pr(-1 \leq \theta \leq 1|Y) = 0.015$

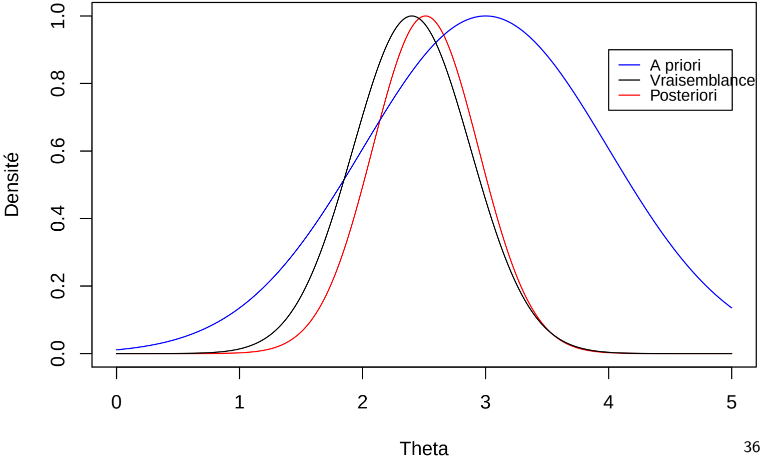
Probabilité d'une hypothèse spécifique



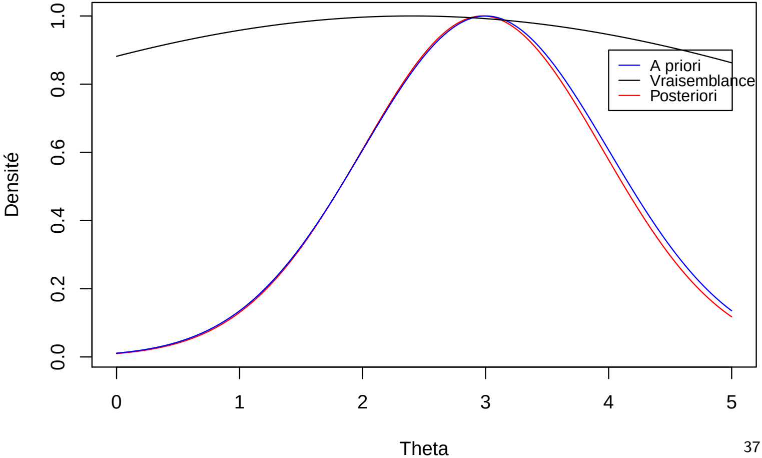
Exemple d'analyse bayésienne

Que se passe-t-il si nous avons les mêmes résultats et la même loi a priori informative $\mathcal{N}(3,1)$, mais avec un $n = 200$... ou un $n = 2$.

c) Approche avec loi a priori informative ($n = 200$)



d) Approche avec loi a priori informative ($n = 2$)



Exemple d'analyse bayésienne

Conclusions selon la taille d'échantillon :

c) Approche avec loi a priori informative ($n = 200$)

La loi a posteriori de θ est dominée par la vraisemblance.
En clair, les approches bayésienne et classique donnent les mêmes résultats pour un grand n!

d) Approche avec loi a priori informative ($n = 2$)

La loi a posteriori de θ est dominée par la loi a priori.
En clair, l'approche bayésienne permet de faire des inférences sur θ même avec de très petits échantillons, ce qui n'est pas le cas pour l'approche classique!

Analyses et logiciels

Analyses accessibles sans programmation en version bayésienne

- corrélation de Pearson, table de contingence
- test t, ANOVA, ANCOVA
- modèles mixtes analyses multiniveaux
- régression linéaire, logistique, de Poisson
- modèles linéaires généralisés
- analyse de survie
- modèle de mélange

Analyses et logiciels

Il existe divers logiciels open-source (STAN, WinBugs, packages R) pour réaliser des analyses bayésiennes simples ou complexes.

Dans les logiciels commerciaux, SAS 9.4, STATA et Mplus 7 permettent ces analyses. SPSS a débuté en 2017 l'introduction de procédures bayésiennes pour quelques analyses spécifiques (test t, corrélation, régression linéaire, ANOVA et régression de Poisson).

Conclusion

Certains puristes voient l'approche bayésienne comme la solution pour remplacer les tests d'hypothèses. Toutefois, la majorité des statisticiens considèrent plutôt cette approche comme un *complément* à l'approche classique.

L'approche bayésienne est particulièrement intéressante pour les situations suivantes :

- Intégrer l'information disponible a priori sur le paramètre d'intérêt
- Analyser un nombre limité d'observations
- Explorer un effet "non-significatif"
- Mettre à jour régulièrement l'inférence sur le paramètre (p.ex., analyses intérimaires pour un essai clinique randomisé)

Références

- Berry, D.A. (2006). Bayesian clinical trials. *Nature Reviews, 5*, 27-36. [intro]
- Chamberlain, D.B., & Chamberlain, J.M. (2017). Making sense of a negative clinical trial result : a Bayesian analysis of a clinical trial of Lorazepam and diazepam for pediatric status epilepticus. *Annals of Emergency Medicine, 69(1)*, 117-124.[illustration pour RCT négatif]
- Spiegelhalter, D.J., Abrams, K.R., & Myles, J.P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley. [contenu peu maths, dispo PDF]
- Wasserstein, R.L., & Lazar, N.A. (2016). The ASA's statement on p-values : context, process and purpose. *The American Statistician, 70(2)*, 129-133. [a lire!!!]