Analyse bayésienne des modèles de médiation et de modération

J-M. GALHARRET, A. PHILIPPE

Laboratoire de Mathématiques Jean Leray

07 Juin 2021

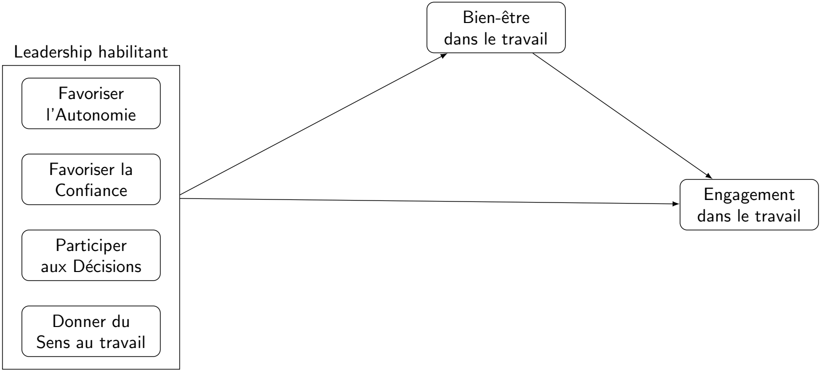
Etude de l'influence du leadership habilitant sur le bien-être et sur l'engagement dans le travail.

- A. Caillé, N. Courtois, J.-M. Galharret, C. Jeoffrion (2020) Influence du leadership habilitant sur le bien-être au travail et l'engagement organisationnel.[Psychologie du travail et des organisations]
- Les données ont été recueillies en 2018 sur un échantillon de $N = 255$ employés de l'industrie aéronautique et $N = 173$ sapeurs pompiers.
- Les variables ont été mesurées à partir de questionnaires en passation collective.

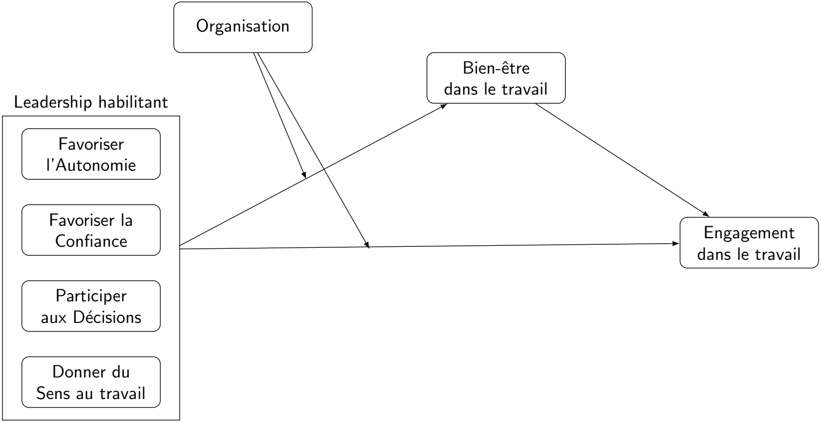
Variables	LH[1]	BEPT	EOA
Outil	Ahearne et al. (2005)	Gilbert et al. (2011)	Meyer et al. (1993)
Nb Items	12	22	6
Cotation		1 à 5	

[1] Constitué de 4 variables : Autonomie, Confiance, Décision, Sens

Médiation

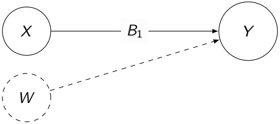

Leadership habilitant
Favoriser l'Autonomie
Favoriser la Confiance
Participer aux Décisions
Donner du Sens au travail
Bien-être dans le travail
Engagement dans le travail

Médiation modérée



Effets - Modèle de médiation dans le cas gaussien

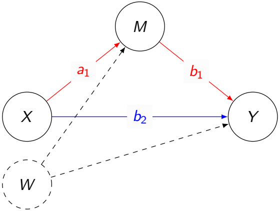
Régression : $X \in \mathbb{R}^k$ explicative d'intérêt, $W \in \mathbb{R}^q$ covariables et $Y \in \mathbb{R}$ réponse.



$$Y = B_0 + B_1 X + B_2 W + \varepsilon$$

Effet de X sur Y : B_1

Médiation : $M \in \mathbb{R}$ médiateur.



$$\begin{cases} M &= a_0 + a_1 X + a_2 W + \varepsilon_M \\ Y &= b_0 + b_1 M + b_2 X + b_3 W + \varepsilon_Y \end{cases}$$

$$\rightsquigarrow Y = b_0 + a_0 b_1 + ({\color{blue}b_2} + {\color{red}a_1 b_1})X + (b_3 + a_2 b_2)W + b_2 \varepsilon_M + \varepsilon_Y,$$

Décomposition du lien entre X et Y :

$$\underbrace{B_1}_{\text{Total}} = \underbrace{{\color{blue}b_2}}_{\text{Direct}} + \underbrace{{\color{red}a_1 b_1}}_{\text{Indirect}}.$$

Inférence sur les effets

- **Approche par M-V :**(Baron and Kenny, 1986) :
  - MV pour l'estimation ponctuelle.
  - Δ-Méthode pour l'intervalle de confiance de l'effet indirect (Sobel, 1982).

- **Approche alternative pour $a_1 b_1$:** MacKinnon et al. (2004) proposent d'estimer par bootstrap la loi de l'estimateur du produit $a_1 b_1$.

- **Une approche bayésienne :** Yuan and MacKinnon (2009) proposent une approche bayésienne avec des lois a priori gaussiennes indépendantes.

Définition du modèle

On considère le modèle de médiation

$$\begin{cases} M &= a_0 + a_1 X + a_2 W + \varepsilon_M, \qquad \alpha = (a_0, a_1, a_2) \\ Y &= b_0 + b_1 M + b_2 X + b_3 W + \varepsilon_Y, \quad \beta = (b_0, b_1, b_2, b_3) \end{cases}$$

On suppose que : $\varepsilon_M, \varepsilon_Y$ variables indépendantes gaussiennes centrées de variances σ_M^2, σ_Y^2.

1. La vraisemblance du modèle de médiation s'écrit :

   $$f(Y, M | \alpha, \beta, \sigma_M^2, \sigma_Y^2, X, W) = \phi_1(Y | \beta, \sigma_Y^2, M, X, W)\phi_2(M | \alpha, \sigma_M^2, X, W)$$

   où ϕ_1, ϕ_2 sont des densités gaussiennes .
2. Choix de la loi a priori pour $\theta = (\alpha, \sigma_M^2, \beta, \sigma_Y^2)$

G-priors pour un modèle de régression Zellner (1971)

Soit un modèle de régression

$$Y = \mathbf{X}\beta + \varepsilon$$

où $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n), \mathbf{X} = [\mathbb{1}, X_1, ..., X_p]$.

Le G-prior des paramètres β, σ^2 est défini par

$$\begin{cases} \beta | \sigma^2, \mathbf{X} \sim \mathcal{N}_{p+1}\left(\widetilde{\beta}, g\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right) \\ \pi(\sigma^2|\mathbf{X}) \propto \sigma^{-2} \end{cases}$$

Choix des hyperparamètres $(g, \widetilde{\beta})$

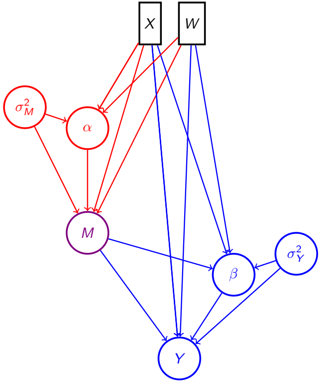
- $\widetilde{\beta} = 0$ et $g = n$ donnent à la loi a priori le même poids qu'une observation. $\rightsquigarrow$ loi peu informative.

Adaptation au modèle de médiation

Décomposition de la loi jointe du modèle :

$$f(M, Y, \alpha, \sigma_M^2, \beta, \sigma_Y^2 | X, W) = \phi_1(Y|\beta, \sigma_Y^2, X, M, W)\pi(\beta, \sigma_Y^2 | X, M, W) \times \phi_2(M|\alpha, \sigma_M^2, X, W)\pi(\alpha, \sigma_M^2 | X, W).$$

- $\pi(\beta, \sigma_Y^2 | X, M, W)$ le $G-$prior de $Y = b_0 + b_1 M + b_2 X + b_3 W + \varepsilon_Y$,
- $\pi(\alpha, \sigma_M^2 | X, W)$ le $G-$prior de $M = a_0 + a_1 X + a_2 W + \varepsilon_M$.



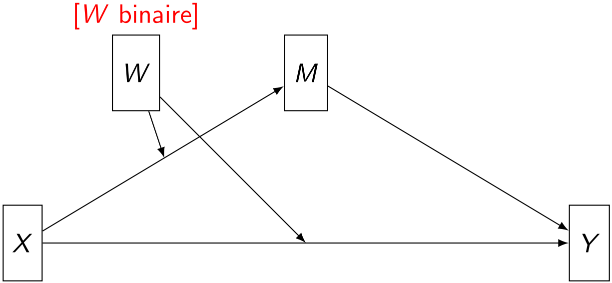
Application aux données sur l'influence du LH

- Dans l'étude, approche classique.
- Résultats sont similaires (pas d'information a priori).
- Inclure de l'information $\rightsquigarrow$ gain en terme de précision.

$n = 428$		Bayésien			Fréquentiste		
			I.Créd.(95%)			I.Conf.(95%)	
		estimate	Lower	Upper	estimate	Lower	Upper
Directs b_2	Sens	0.161	0.073	0.247	0.161	0.036	0.279
	Décision	0.111	0.031	0.193	0.112	0.014	0.208
	Confiance	0.048	-0.050	0.150	0.049	-0.077	0.172
	Autonomie	0.053	-0.043	0.148	0.051	-0.069	0.169
Indirects $a_1 b_1$	Sens	0.059	0.025	0.093	0.060	0.032	0.101
	Décision	0.041	-0.034	0.009	-0.012	-0.036	0.006
	Confiance	0.017	0.008	0.068	0.036	0.008	0.076
	Autonomie	0.019	-0.016	0.033	0.009	-0.016	0.038

Modèle de médiation modérée

Pour comparer les effets dans deux populations différentes on ajoute des termes d'interaction dans les équations de régression.



$$[\mathcal{M}_1] : \begin{cases} M = a_0 + a_1 X + a_2 W + a_3 \; X : W + \varepsilon_M, \\ Y = b_0 + b_1 M + b_2 X + b_3 W + b_4 \; X : W + \varepsilon_Y \end{cases}$$

$\mathcal{H}_0 : a_3 = b_4 = 0$ contre $\mathcal{H}_1 : a_3 \neq 0$ ou $b_4 \neq 0$.

Test et sélection de modèle (par exemple Marin and Robert, 2014)

On exprime le test comme un problème de sélection de modèle. Le test

$$(\theta_0, \theta_1) \in \Theta_0 \times \{0\} \text{ contre } (\theta_0, \theta_1) \in \Theta_0 \times \Theta_1$$

devient un problème de choix de modèle entre

$$\mathcal{M}_0 = \{f_0(\mathcal{D}|\theta_0), \mathsf{supp}(\pi_0) = \Theta_0\}$$

$$\mathcal{M}_1 = \{f_1(\mathcal{D}|\theta_0, \theta_1)); \mathsf{supp}(\pi_1) = \Theta_0 \times \Theta_1\}$$

Facteur de Bayes

Le facteur de Bayes de $\mathcal{M}_1$ contre $\mathcal{M}_0$

$$BF_{10} = \frac{\mathfrak{M}^1(\mathcal{D})}{\mathfrak{M}^0(\mathcal{D})} \text{ où } \mathfrak{M}^i(\mathcal{D}) = \int_{\Theta_i} f_i(\mathcal{D}|\theta_i)\pi_i(\theta_i)\mathrm{d}\theta_i.$$

- Règle de décision :

  On choisit $\mathcal{M}_1$ lorsque $BF_{10} > 1$.

- Avantage : le rôle des hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$ est symétrique.
- Inconvénient : $\rightsquigarrow$ Pas de niveau de confiance.
- Solution : échelles d'évidence (Jeffreys, 1961; Kass and Raftery, 1995).

$\log_{10}(BF_{10})$	$[0, 0.5[$	$[0.5, 1[$	$[1, 2[$	$[2, +\infty[$
certitude que $\mathcal{H}_0$ fausse	faible	substantielle	forte	décisive

Facteur de Bayes pour le modèle de régression

- Modèle de régression :

$$\begin{cases} Y = \mathbf{X}\beta + \varepsilon, & \text{où } \beta \in \mathbb{R}^{p+1}. \\ (\beta, \sigma^2) \sim G - \text{prior} \end{cases}$$

- Pour tester

$$\mathcal{H}_0 : \beta_1 = ... = \beta_q = 0, \qquad q < p$$

on compare les modèles de régression :

$$\mathbf{X} = [\mathbb{1}, X_1, ..., X_q, X_{q+1}, ..., X_p] \text{ et } \mathbf{X}_0 = [\mathbb{1}, X_{q+1}, ..., X_p].$$

- Le facteur de Bayes est explicite (Marin and Robert, 2014)

$$BF_{10} = (n+1)^{-q/2} \left(\frac{y'y - \frac{n}{n+1} y'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0 y}{y'y - \frac{n}{n+1} y'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X} y} \right)^{n/2}$$

Application à la médiation modérée

On considère les deux modèles de médiation.

$$[\mathcal{M}_0] : \begin{cases} M = a_0 + a_1 X + a_2 W + \varepsilon_M, \\ Y = b_0 + b_1 M + b_2 X + b_3 W + \varepsilon_Y \end{cases}$$

$$[\mathcal{M}_1] : \begin{cases} M = a_0 + a_1 X + a_2 W + a_3 X : W + \varepsilon_M, \\ Y = b_0 + b_1 M + b_2 X + b_3 W + b_4 X : W + \varepsilon_Y \end{cases}$$

Proposition

Pour un modèle de médiation modérée dans un cadre gaussien dont les G-priors sont définis précédemment, le facteur de Bayes est explicite

$$BF_{10} = BF_{10}^M BF_{10}^Y$$

Application aux données :

- $\Lambda_n := \log(BF_{10}) = -8.635 \rightsquigarrow$ Décisif en faveur de $\mathcal{H}_0$ pour l'échelle de Kass and Raftery (1995) et Jeffreys (1961).
- Test de rapport de vraisemblance : $p = .058$ (Caillé et al. (2020))

Test fréquentiste

On utilise le BF_{10} comme statistique de test fréquentiste.
$\rightsquigarrow$ calcul de la $p - value$ associée à la région critique

$$\{\Lambda_n := \log_{10}(BF_{10}) > q_\alpha\}$$

où q_α est le $(1-\alpha)$-quantile de Λ_n sous $\mathcal{H}_0$.

Exemple :

Modèle de régression : test asymptotique par Zhou and Guan (2018).

Mise en œuvre du test pour la médiation modérée

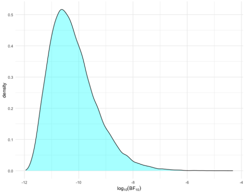
On peut appliquer le bootstrap paramétrique pour approcher le quantile q_α de Λ_n sous $\mathcal{H}_0$. -¿ à dire au début à l'oral.

- On teste $\mathcal{H}_0 : a_3 = b_4 = 0$ contre $\mathcal{H}_1 : a_3 \neq 0$ ou $b_4 \neq 0$.
- Les paramètres du modèle :

$$\begin{cases} (\theta_{nuis}, 0, 0) & \text{sous } \mathcal{H}_0, [\mathcal{M}_0] \\ (\theta_{nuis}, a_3, b_4) & \text{sous } \mathcal{H}_1, [\mathcal{M}_1] \end{cases}$$

- On estime les paramètres du modèle $[\mathcal{M}_1]$: $(\widehat{\theta}_{nuis}, \widehat{a}_3, \widehat{b}_4)$.
- $\widehat{\theta}_{nuis}$ est un estimateur consistant de θ_{nuis} sous $\mathcal{H}_0$ et $\mathcal{H}_1$ (car $\mathcal{M}_0 \subset \mathcal{M}_1$).

- Distribution approchée de Λ_n sous $\mathcal{H}_0 \rightarrow$
- p-value : $\mathbb{P}(\Lambda_n > -8.635) = .055$
- En accord avec le rapport de vraisemblance.



Version bayésienne

- Ayant observé (Y, M), la loi prédictive sous $\mathcal{M}_1$ est

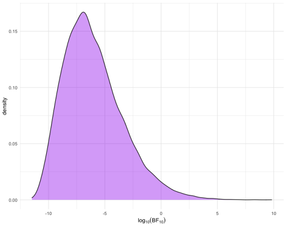
$$p_1(y^\star, m^\star | Y, M) = \int f_1(y^\star, m^\star | Y, M, \theta)\pi_1(\theta | Y, M)\mathrm{d}\theta,$$

  où f_1 est la vraisemblance conditionnelle et π_1 la loi a posteriori sous $\mathcal{M}_1$.
- On calcule la loi du facteur de Bayes :

$$BF_{10}(Y^\star, M^\star), \qquad (Y^\star, M^\star) \sim p_1$$

- On quantifie l'évidence de $\mathcal{M}_1$ contre $\mathcal{M}_0$ par

$$\mathbb{P}(\Lambda_n(M^\star, Y^\star) > 0 | M, Y)$$



On a une faible évidence en faveur de $\mathcal{M}_1$:

$$\mathbb{P}(\Lambda_n(M^\star, Y^\star) > 0 | M, Y) = 0.04.$$

Conclusion

- Les estimations issues de la modélisation bayésienne apportent des résultats similaires à la méthode classique. L'avantage est l'interprétation et/ou l'intégration de données historiques.

- Pour l'effet indirect, on a facilement accès à la loi a posteriori et à des intervalles de crédibilité.

- Du point de vue bayésien, pour tester les effets, on peut utiliser les approches basés sur le BF.

Ahearne, M., Mathieu, J., and Rapp, A. (2005). To empower or not to empower your sales force? An empirical examination of the influence of leadership empowerment behavior on customer satisfaction and performance. *Journal of Applied Psychology*.

Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173.

Caillé, A., Courtois, N., Galharret, J.-M., and Jeoffrion, C. (2020). Influence du leadership habilitant sur le bien-être au travail et l'engagement organisationnel : étude comparative entre une organisation habilitante et une organisation classique. *Psychologie du Travail et des Organisations*, 26(3):247–261.

Gilbert, M.-H., Dagenais-Desmarais, V., and Savoie, A. (2011). Validation d'une mesure de santé psychologique au travail. *European Review of Applied Psychology*, 61(4):195 – 203.

Jeffreys, H. (1961). *Theory of Probbility*. Oxford University Press.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

MacKinnon, D. P., Lockwood, C. M., and Williams, J. (2004).

Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39.

Marin, J.-M. and Robert, C. (2014). *Bayesian essentials with R*. Springer Textbooks in Statistics. Springer Verlag, New York.

Meyer, J. P., Allen, N. J., and Smith, C. A. (1993). Commitment to organizations and occupations: Extension and test of a three-component conceptualization. *Journal of Applied Psychology*, 78(4):538–551.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13.

Yuan, Y. and MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, 14.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. A Wiley-Interscience publication. Wiley.

Zhou, Q. and Guan, Y. (2018). On the null distribution of bayes factor in linear regression. *Journal of the American Statistical Association*, 113(523):1362–1371.