

Combinaison de modèles hétéroscédastiques dans un cadre de provisionnement des pertes

Sébastien Jessup

Concordia University

28 juin 2023

Table des matières

- 1 Introduction
- 2 Combinaison de modèles
 - Combinaison linéaire
 - Bayesian Model Averaging
 - Une nouvelle approche
 - Exemple simplifié
 - Deuxième exemple simplifié
- 3 Application sur des données
 - Données et modèles choisis
 - Résultats

Introduction

Dans tout problème de prévision, on doit modéliser des données.



Modèles simples

- Facilement explicables
- Faciles à utiliser
- Densité connue
- Parfois incomplets

Modèles plus complexes

- Peuvent mieux prédire
- Parfois des boîtes noires
- Densité moins facile à évaluer

Doit-on en choisir un ? Tous les garder ?

Combinaison linéaire

$$f(y) = \sum_{m=1}^M w_m f_m(y),$$

$$\sum_{m=1}^M w_m = 1; 0 \leq w_m \leq 1$$

- L'approche classique : $w_{m^*} = 1$ pour le meilleur modèle m^* .
- Une autre approche : $w_m = 1/M \forall m$.
- Finalement, on peut chercher à optimiser les poids. Comment ?

Bayesian Model Averaging (BMA)

- Avec le théorème de Bayes, on a

$$w_m = \Pr(\mathcal{M}_m | \mathcal{D}) = \frac{\Pr(\mathcal{D} | \mathcal{M}_m) \Pr(\mathcal{M}_m)}{\sum_{l=1}^M \Pr(\mathcal{D} | \mathcal{M}_l) \Pr(\mathcal{M}_l)}$$

- Habituellement, on utilise un algorithme de maximisation des attentes (Raftery et al., 2005) pour obtenir

$$\hat{w} = \arg \max_w \sum_{y \in \mathcal{D}} \log \left(\sum_{m=1}^M w_m f_m(y | \mathcal{M}_m, \sigma_m^2) \right),$$

où on met à jour les poids en supposant une variance homoscédastique.

BMA hétéroscédastique

- Remarquez que la formule précédente optimisait une variance globale pour chaque modèle m .
- Très souvent, l'hypothèse d'homoscédasticité ne tient pas.
- Il faudrait alors adapter l'algorithme pour considérer l'hétéroscédasticité.

BMA hétéroscédastique (suite)

- On veut évaluer $y^{(k)}$, la valeur pour la $k^{(e)}$ observation.
- On peut supposer que $Y^{(k)} - E(\hat{Y}_m^{(k)}) \sim N(0, \sigma_k^2)$.
- On considère alors

$$\hat{w} = \arg \max_w \sum_{y^{(k)} \in \mathcal{D}} \log \left(\sum_{m=1}^M w_m f_m^*(y^{(k)}) \right),$$

avec f_m^* une distribution normale de moyenne $E(\hat{Y}_m^{(k)})$ et variance $\sigma_{k,m}^2$.

Mise à jour avec hétéroscédasticité

- On obtient la proportion à chaque modèle telle que

$$z_{m,k}^{(j)} = \frac{w_m^{(j-1)} \phi \left(\frac{y^{(k)} - E(\hat{Y}_m^{(k)})}{\sigma_{k,m}} \right)}{\sum_{m=1}^M w_m^{(j-1)} \phi \left(\frac{y^{(k)} - E(\hat{Y}_m^{(k)})}{\sigma_{k,m}} \right)}.$$

- On met à jour les poids en prenant la moyenne :

$$w_m^{(j)} = \frac{1}{K} \sum_{k=1}^K z_{m,k}^{(j)}$$

Le problème de convergence à 1

- L'algorithme de maximisation de BMA a le problème connu de converger vers un seul modèle.
- La solution pratique est généralement d'arrêter l'algorithme après un certain nombre d'itérations.
- La convergence pourrait être due au fait qu'on considère les données observées comme les seules données possibles.
- On a $Y^{(k)} \sim \mathbb{P}$ et $\hat{Y}_m^{(k)} \sim \mathbb{Q}_m$, pour des distributions \mathbb{P} et \mathbb{Q}_m .
- On veut comparer \mathbb{Q}_m avec \mathbb{P} , pas la réalisation $y^{(k)}$.

Une nouvelle approche

- Supposons que $Y^{(k)}$ peut être raisonnablement bien approximée par $N(E(Y^{(k)}), \sigma_k^2)$.
- Alors $y^{(k)} = E(Y^{(k)}) + \epsilon_k$, où $\epsilon_k \sim N(0, \sigma_k^2)$.
- Supposons également que $E(\hat{Y}_m^{(k)}) = E(Y^{(k)})$.
- Alors $y^{(k)} - E(\hat{Y}_m^{(k)}) \sim N(0, \sigma_k^2)$.

Intégrer pour sortir l'erreur aléatoire

$$\begin{aligned}\Pr(\mathcal{M}_m|\mathcal{D}) &= \int \Pr(\mathcal{M}_m|\mathcal{D}, \epsilon) \pi(\epsilon|\mathcal{D}) d\epsilon \\ &= \int \cdots \int \frac{\Pr(\mathcal{D}|\mathcal{M}_m, \epsilon) \Pr(\mathcal{M}_m)}{\sum_{l=1}^M \Pr(\mathcal{D}|\mathcal{M}_l, \epsilon) \Pr(\mathcal{M}_l)} \times \\ &\quad \pi(\epsilon_m^{(1)}|\mathcal{D}) \cdots \pi(\epsilon_m^{(K)}|\mathcal{D}) d\epsilon^{(1)} \cdots d\epsilon^{(K)}\end{aligned}$$

- Avec S simulations, on obtient

$$\approx \frac{1}{S} \sum_{s=1}^S \frac{1}{K} \sum_{k=1}^K \frac{\Pr(y^{(k)}|\epsilon_{m,s}^{(k)}, \mathcal{M}_m)}{\sum_{l=1}^M \Pr(y^{(k)}|\epsilon_{l,s}^{(k)}, \mathcal{M}_l)}.$$

- Par contre, ceci demanderait de connaître la variance σ_k^2 de l'erreur aléatoire.

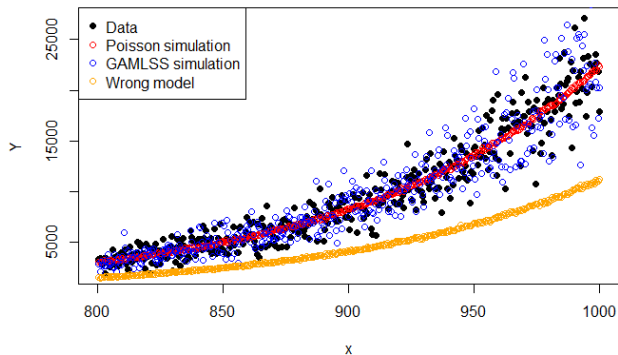
Estimer l'erreur σ_k^2

- On peut estimer l'erreur en supposant que $E(\sigma_k^2) = e^{X_k\beta}$ selon l'idée de Harvey (1976).
- En définissant $R_{m,k} = \ln((y^{(k)} - E(\hat{Y}_m^{(k)}))^2)$, on a

$$\begin{aligned}\hat{\beta}_m &= \arg \min_{\beta_m} \sum_{k=1}^K \left(\ln((y^{(k)} - E(\hat{Y}_m^{(k)}))^2) - X^{(k)}\beta_m \right)^2 \\ &= (X^T X)^{-1} X^T R_m.\end{aligned}$$

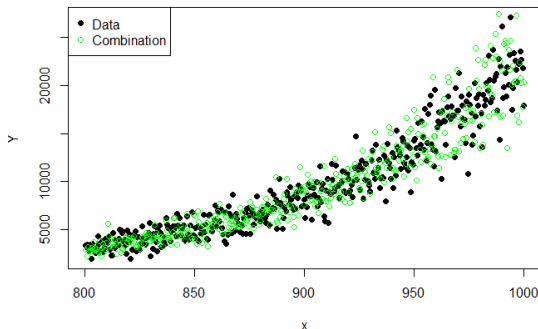
- On peut alors utiliser l'intégrale numérique pour obtenir des poids considérant \mathbb{P} plutôt qu'une seule réalisation.

Exemple simplifié



Exemple simplifié - suite

- 100% du poids sur le modèle GAMLSS avec l'algorithme classique, contre 88.6% avec l'intégrale numérique.
- Le mauvais modèle reçoit seulement 0.8% du poids avec l'intégrale numérique.



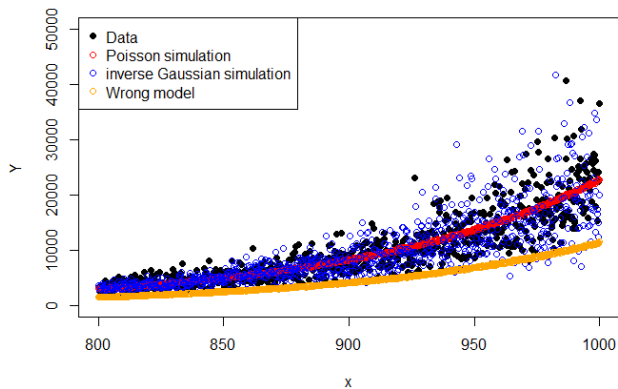
La problématique de l'incertitude non-symétrique

- Il est possible que l'erreur ne soit pas symétrique.
- On doit alors essayer de modéliser l'asymétrie pour avoir $E(\gamma) = g(X_k \zeta_k)$, tel que

$$\hat{\zeta}_m = \arg \min_{\zeta_m} \sum_{k=1}^K \left(g \left(\frac{(y^{(k)} - E(\hat{Y}_m^{(k)}))^3}{\tilde{\sigma}_{m,k}^3} \right) - X^{(k)} \zeta_m \right)^2.$$

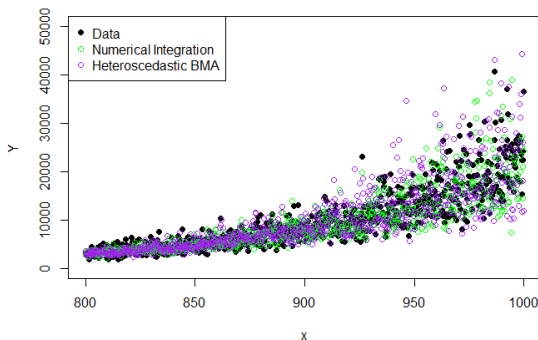
- En séparant les termes positifs et négatifs, on peut utiliser \ln comme fonction.
- Selon l'asymétrie observée, on peut alors utiliser une "skew normal" ou une distribution GEV.

Deuxième exemple simplifié



Deuxième exemple simplifié - suite

- 100% du poids sur le modèle inverse gaussien avec l'algorithme classique, contre 86.7% avec l'intégration d'erreur.
- Le mauvais modèle reçoit cette fois-ci 3.2%.



Données

- On travaille avec une base de données simulée à partir du "simulation machine" de Gabrielli & Wuthrich (2018).

CINr	LoB	AY	AQ	age	cc	inj_part	RepDel	CY	Paid	St	ult_paid
1	1	1994	2	33	45	20	0	1994	63	0	63
2	1	1994	2	31	13	20	0	1994	0	0	171
3	1	1994	1	39	2	30	0	1994	0	0	289
4	1	1994	4	68	44	18	1	1994	0	0	224
5	1	1994	2	41	18	20	0	1994	0	0	618
6	1	1994	3	43	44	35	0	1994	0	1	14102

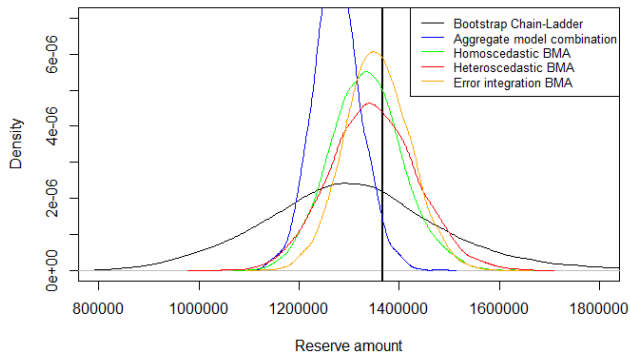
Modèles choisis

- Avanzi et al. (2022) ont établi des critères désirables pour combiner des modèles :
 - modèles qui peuvent être ajustés automatiquement,
 - modèles avec différentes forces et limitations,
 - modèles qui sont facilement identifiables et interprétables.
- Pour le moment, on travaille avec des GLMs granulaires, un DGLM granulaire, et des modèles agrégés appliqués aux données individuelles.

Résultats

Modèle	Poids (%) du BMA basé sur		
	Homoscédastique	Hétéroscédastique	Intégration numérique
GLM Gamma	13.6	0.4	18.0
GLM ODP	10.0	0.5	17.8
DGLM ODP	8.5	98.4	26.5
GLM agrégé	33.7	0.5	17.4
GAMLSS agrégé	34.2	0.2	20.3

Résultats (suite)



Résumé et prochaines étapes

- On compare donc principalement deux méthodes :
 - un algorithme de BMA hétéroscédastique,
 - et un algorithme d'intégration numérique d'erreur.
- Jusqu'à maintenant il y a un seul article sur la combinaison de modèles agrégés, et rien sur la combinaison de modèles granulaires.
- Notre algorithme est capable d'automatiquement prendre en compte des cas où l'incertitude est asymétrique, et règle le problème de convergence à un seul modèle.
- Vos suggestions sont les bienvenues, merci d'avoir écouté!

Bibliographie

Avanzi, B., Li, Y., Wong, B., & Xian, A. (2022). Ensemble distributional forecasting for insurance loss reserving. arXiv preprint arXiv :2206.08541.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica : Journal of the Econometric Society*, 461-465.

Gabrielli, A., & Wüthrich, M. (2018). An individual claims history simulation machine. *Risks*, 6(2), 29.

Kapetanios, G., Mitchell, J., Price, S., & Fawcett, N. (2015). Generalised density forecast combinations. *Journal of Econometrics*, 188(1), 150-165.