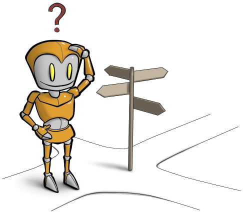
Combien serons-nous en France en 2070 ?

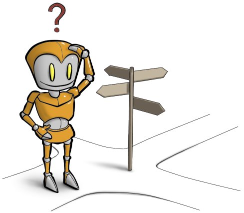
Une réponse à l'aide de projections probabilistes bayésiennes

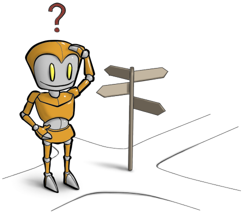
Vianney Costemalle

Insee, DMS, DMRG

14 juin 2018
Journées de Méthodologie Statistique





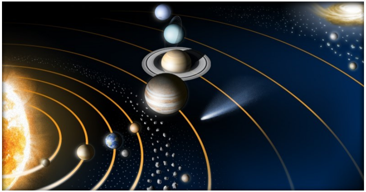



A
K
Q
J
10

Données $\rightarrow$ **Evolution (θ)** $\rightarrow$ Projections

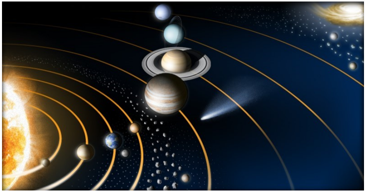
Données $\rightarrow$ **Evolution (θ)** $\rightarrow$ Projections

θ fixé $\Rightarrow$ **déterministe**



Données $\rightarrow$ **Evolution (θ)** $\rightarrow$ Projections

θ fixé $\Rightarrow$ **déterministe**



θ aléatoire $\Rightarrow$ **probabiliste**

Projections probabilistes

- **stochastiques** : termes d'erreur ϵ
- **bayésiennes** : on considère les paramètres θ comme des variables aléatoires

$$p(\theta|donnees) = \frac{p(donnees|\theta).p(\theta)}{p(donnees)}$$

distribution *a priori* $\rightarrow$ distribution *a posteriori*

Cohort-component method

décomposition par **groupes d'âge**
3 composantes : **fécondité**, **mortalité**, **migrations**

Notations

- **P(a,n,s)** : nbr de personnes au 1er janvier de l'année n, de sexe s nées l'année $n-a$
  $P(0,n,s)$:nbr de naissances vivantes l'année n de bébés de sexe s
- **D(a,n,s)** : nbr de décès durant l'année n, de personnes nées l'année $n-a$ et de sexe s
- **N(a,n,s)** : nbr de bébés de sexe s nés vivants durant l'année n et dont la mère est née l'année $n-a$
- **M(a,n,s)** : solde migratoire de l'année n, pour les personnes de sexe s et nées l'année $n-a$

$P(n) = \sum_{a \geq 1,s} P(a,n,s)$
$M(n) = \sum_{a,s} M(a,n,s)$
$N(a,n) = N(a,n,filles) + N(a,n,garçons)$

Cohort-component method

décomposition par **groupes d'âge**
3 composantes : **fécondité**, **mortalité**, **migrations**

Notations

- **P(a,n,s)** : nbr de personnes au 1er janvier de l'année n, de sexe s nées l'année $n-a$
  $P(0,n,s)$:nbr de naissances vivantes l'année n de bébés de sexe s
- **D(a,n,s)** : nbr de décès durant l'année n, de personnes nées l'année $n-a$ et de sexe s
- **N(a,n,s)** : nbr de bébés de sexe s nés vivants durant l'année n et dont la mère est née l'année $n-a$
- **M(a,n,s)** : solde migratoire de l'année n, pour les personnes de sexe s et nées l'année $n-a$

$P(n) = \sum_{a\geq 1,s} P(a,n,s)$
$M(n) = \sum_{a,s} M(a,n,s)$
$N(a,n) = N(a,n,filles) + N(a,n,garçons)$

Cohort-component method

Populations à "risque"

On raisonne en "personnes-années"

Pour les décès :

$R_D(a, n, s) = P(a, n, s) + \frac{1}{2}M(a, n, s)$ si $a \geq 1$

$R_D(0, n, s) = \frac{1}{2}P(0, n, s) + \frac{1}{2}M(a, n, s)$ si $a = 0$

Pour les naissances :

$R_N(a, n) = P(a, n, femmes) + \frac{1}{2}M(a, n, femmes) - \frac{1}{2}D(a, n, femmes)$

Cohort-component method

Evolution de la population, d'une période à une autre (période = 1 an)

$$P(a, n, s) = P(a-1, n-1, s) + M(a-1, n-1, s) - D(a-1, n-1, s)$$

pour $a \geq 1$ et

$$P(0, n, s) = N(n, s) = \sum_a N(a, n, s)$$

Il ne reste plus qu'à ...
déterminer $M(a, n, s)$, $D(a, n, s)$ et $N(a, n, s)$ pour
$a \in \{0, ..., 100\}$
$n \in \{2014, ..., 2070\}$
$s \in \{femmes, hommes\}$

... soit environ 35 000 valeurs

Cohort-component method

Evolution de la population, d'une période à une autre (période = 1 an)

$$P(a,n,s) = P(a-1,n-1,s) + M(a-1,n-1,s) - D(a-1,n-1,s)$$

pour $a \geq 1$ et

$$P(0,n,s) = N(n,s) = \sum_a N(a,n,s)$$

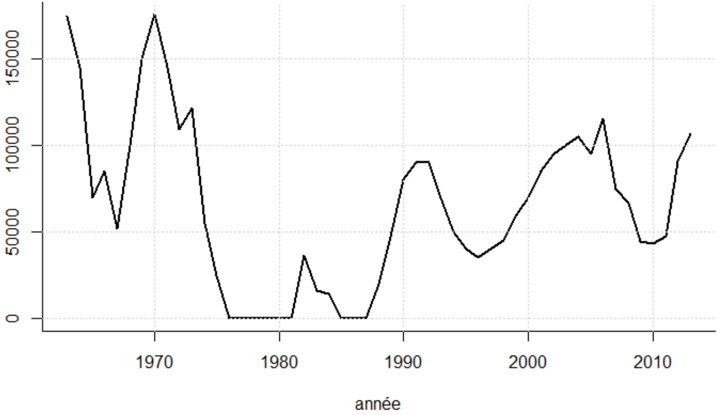
Il ne reste plus qu'à ...
déterminer $M(a,n,s)$, $D(a,n,s)$ et $N(a,n,s)$ pour
$a \in \{0,...,100\}$
$n \in \{2014,...,2070\}$
$s \in \{femmes, hommes\}$

... soit environ 35 000 valeurs

Littérature

- Blanpain et Buisson (2016) "Projections de population 2013-2070 pour la France : méthode et principaux résultats", *document de travail Insee*
- ONU (2017) "World Population Prospects : the 2017 revision, key findings and advance tables", *working paper 248*
- Gerland et al (2014) "World population stabilization unlikely this century" *Science*
- Dunstan et Ball (2016) "Demographic projections : user and producer experiences of adopting a stochastic approach", *Journal of Officials Statistics*
- Bijak et al. (2015) "Probabilistic population forecasts for informed decision making", *Journal of Officials Statistics*
- Lee et Carter (1992) "Modeling and forecasting US mortality", *Journal of the American Statistical Association*

Les migrations - données



Les migrations - modèle

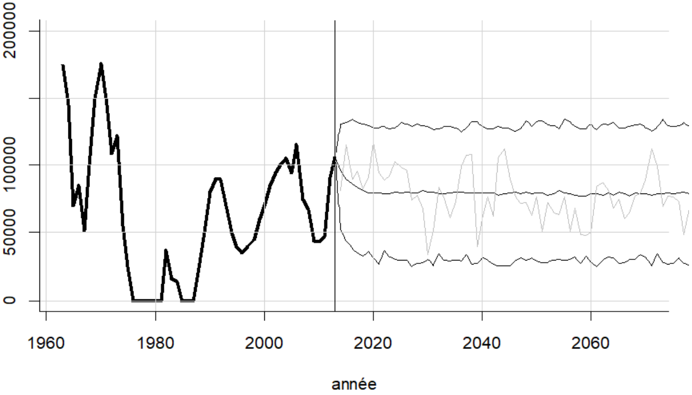
Modèle autorégressif d'ordre 1 :

$$M(n) = M_{lt} + \rho_M(M(n-1) - M_{lt}) + \epsilon_M(n)$$
$$\epsilon_M(n) \sim \mathcal{N}(0, \sigma_M)$$

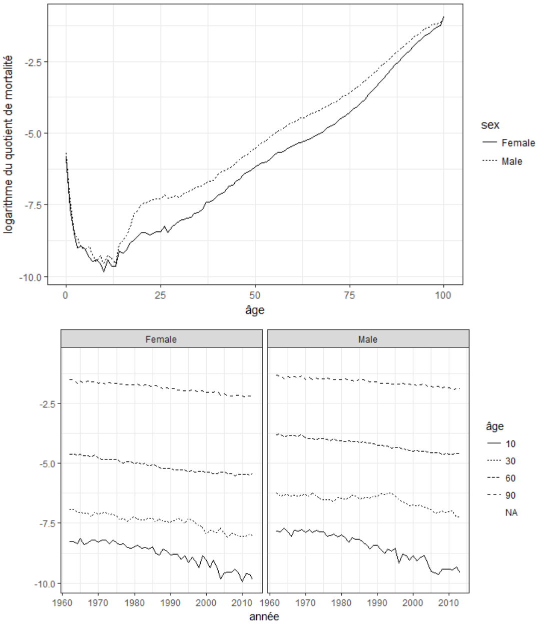
1 - <u>Estimation des lois a posteriori</u>
$M_{lt} \sim \mathcal{N}(80\ 000, 10\ 000)$
$\rho_M \sim \mathcal{B}eta(10, 10)$

2 - <u>Simulation de trajectoires</u>:
a) $(M_{lt}^s, \rho_M^s, \sigma_M^s) \sim \mathcal{L}(.|donnees)$ pour $s \in \{1, ..., 1000\}$
b) on fait évoluer $M(n)$ pour n de 2014 à 2070

Les migrations - résultats



Les décès - données



Les décès - modèle

$$D(a,n,s) \sim \mathcal{P}oisson(\mu_D(a,n,s).R_D(a,n,s))$$
$$log(\mu_D(a,n,s)) = \beta^0 + \beta_a^{age} + \beta_{a,s}^{age:sexe} + \beta_{a,n}^{age:annee} + \epsilon_{D,1}(a,n,s)$$

et

$$\beta_{a,n}^{age:annee} = \theta_{a,n}^{age:annee} + \eta(a,n)$$
$$\theta_{a,n}^{age:annee} = \theta_{a,n-1}^{age:annee} + \delta_{a,n}^{age:annee} + \nu(a,n)$$
$$\delta_{a,n}^{age:annee} = \delta_{a,n-1}^{age:annee} + \omega(a,n)$$

Estimations des lois *a posteriori* avec le **package R** `demest` (John Bryant, Jenny Harlow, Mark Scambary et Junni Zhang - `https://github.com/StatisticsNZ`)

Les décès - modèle

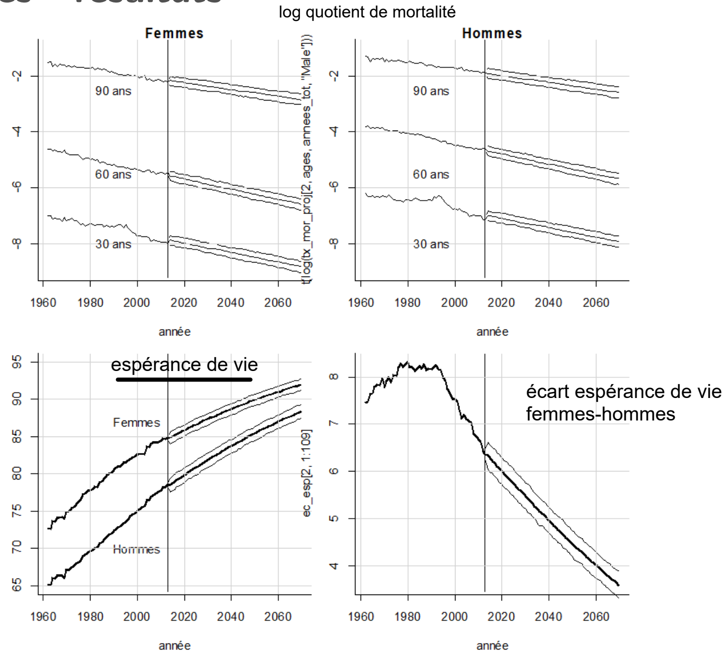
$$D(a,n,s) \sim \mathcal{P}oisson(\mu_D(a,n,s).R_D(a,n,s))$$
$$log(\mu_D(a,n,s)) = \beta^0 + \beta_a^{age} + \beta_{a,s}^{age:sexe} + \beta_{a,n}^{age:annee} + \epsilon_{D,1}(a,n,s)$$

et

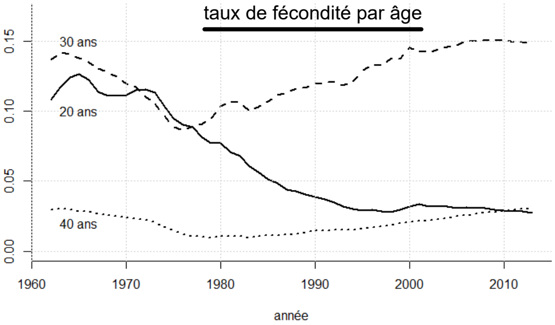
$$\begin{aligned}
\beta_{a,n}^{age:annee} &= \theta_{a,n}^{age:annee} + \eta(a,n) \\
\theta_{a,n}^{age:annee} &= \theta_{a,n-1}^{age:annee} + \delta_{a,n}^{age:annee} + \nu(a,n) \\
\delta_{a,n}^{age:annee} &= \delta_{a,n-1}^{age:annee} + \omega(a,n)
\end{aligned}$$

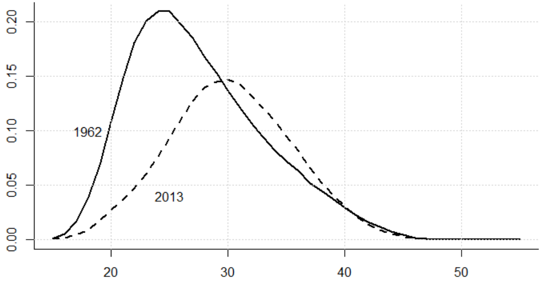
Estimations des lois *a posteriori* avec le **package R** `demest` (John Bryant, Jenny Harlow, Mark Scambary et Junni Zhang - `https://github.com/StatisticsNZ`)

Les décès - résultats

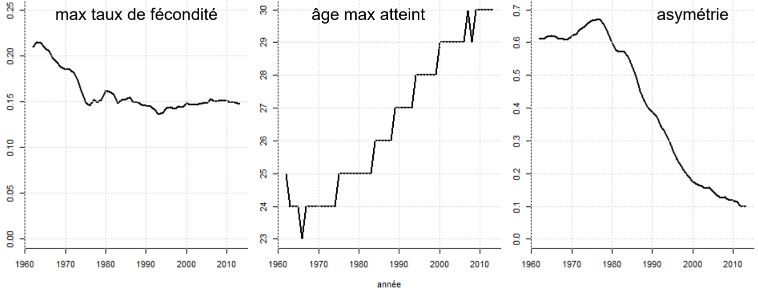


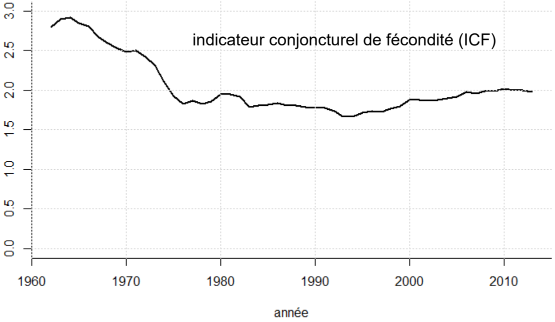
Les naissances - données





Les naissances - données





Les naissances - méthode

1. on projette l'ICF selon un modèle autorégressif d'ordre 1
2. on projette les taux de fécondité par âge (Bijak et al. 2015)

$$N(a,n) \sim \mathcal{P}oisson(\mu_F(a,n).R_F(a,n))$$

$$log(\mu_F(a,n)) = \alpha_a + \beta_a \kappa_n + \gamma_{n-a} + \epsilon_F(a,n)$$

$$\kappa_n = \phi_0 + \phi_1 \kappa_{n-1} + \xi(n)$$
$$\gamma_{n-a} = \psi_0 + \psi_1 \gamma_{n-a-1} + \zeta(n)$$

3. on cale les taux de fécondité pour retomber sur l'ICF projeté

Les naissances - méthode

1. on projette l'ICF selon un modèle autorégressif d'ordre 1

2. on projette les taux de fécondité par âge (Bijak et al. 2015)

$$N(a,n) \sim \mathcal{P}oisson(\mu_F(a,n).R_F(a,n))$$

$$log(\mu_F(a,n)) = \alpha_a + \beta_a \kappa_n + \gamma_{n-a} + \epsilon_F(a,n)$$

$$\kappa_n = \phi_0 + \phi_1 \kappa_{n-1} + \xi(n)$$
$$\gamma_{n-a} = \psi_0 + \psi_1 \gamma_{n-a-1} + \zeta(n)$$

3. on cale les taux de fécondité pour retomber sur l'ICF projeté

Les naissances - méthode

1. on projette l'ICF selon un modèle autorégressif d'ordre 1

2. on projette les taux de fécondité par âge (Bijak et al. 2015)

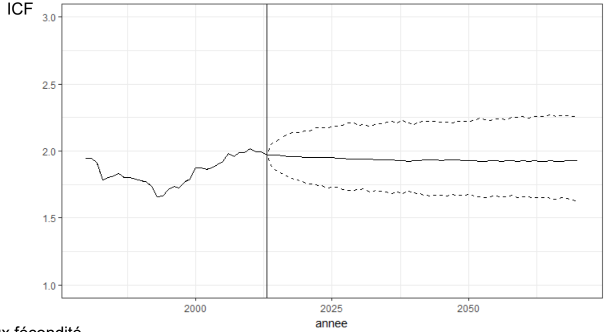
$$N(a,n) \sim \mathcal{P}oisson(\mu_F(a,n).R_F(a,n))$$

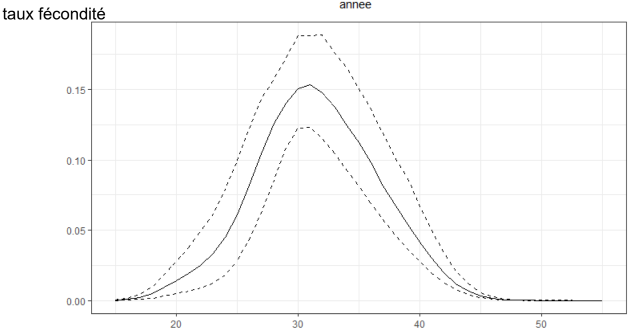
$$log(\mu_F(a,n)) = \alpha_a + \beta_a \kappa_n + \gamma_{n-a} + \epsilon_F(a,n)$$

$$\kappa_n = \phi_0 + \phi_1 \kappa_{n-1} + \xi(n)$$
$$\gamma_{n-a} = \psi_0 + \psi_1 \gamma_{n-a-1} + \zeta(n)$$

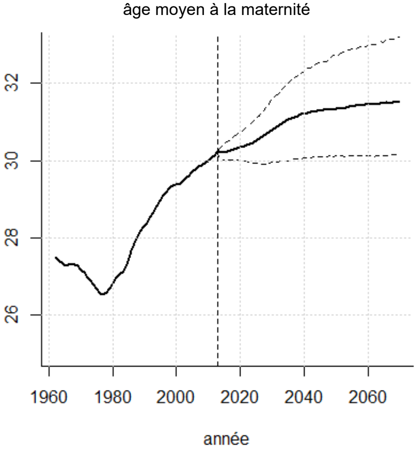
3. on cale les taux de fécondité pour retomber sur l'ICF projeté

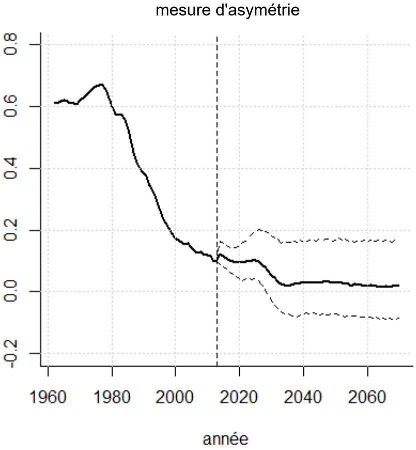
Les naissances - résultats (1/2)



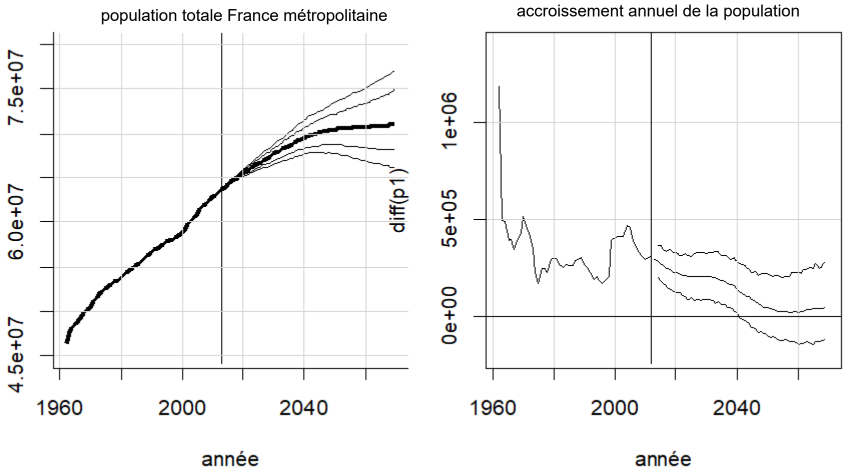


Les naissances - résultats (2/2)

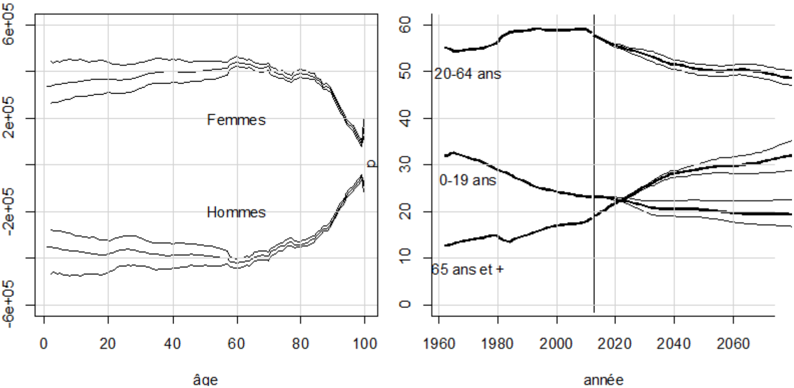




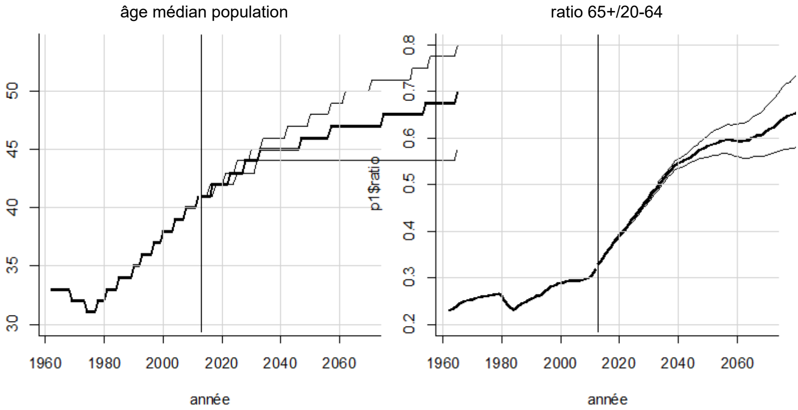
La population - résultats (1/3)



La population - résultats (2/3)



La population - résultats (3/3)



Conclusion

Projections déterministes

+ simple à communiquer
+ scénarios particuliers
+ scénarios extrêmes
- multiplication des sénarios
- scénarios = probabilité nulle

Projections probabilistes bayésiennes

+ quantification de l'incertitude
+ calculer n'importe quel indicateur
+ incorporer des savoirs *a priori*