Muller Timothée

Projet de simulation par chaînes de Markov

26 mai 2014

M. Bérard

Table des matières

Introduction

Tout d'abord, ce fichier sera simplement la justification des différentes formules et méthodes théoriques utilisées pour ce projet. Chaque chapitre de ce fichier concerne sa partie respectivement numérotée dans le code. L'ensemble des informations concernant les fonctions se trouvent en commentaire au sein du code.

Ensuite, le but de ce projet a été d'estimer la prime pure d'un contrat d'assurance hospitalisation pour des femmes âgées de 20 à 65 ans. Deux principales hypothèses ont été faites :

- il n'y a que 3 états possibles :
  - 1 = actif (non hospitalisé)
  - 2 = hospitalisation (hors grossesse)
  - 3 = hospitalisation pour grossesse
  - le décès n'interviendra donc nulle part tout au long de ce projet
- le passage d'un état d'hospitalisation à l'autre n'est possible qu'avec transition obligatoire par l'état actif ce que l'on peut traduire de la façon suivante :

$$p_{2->3} = 0$$
$$p_{3->2} = 0$$

Les différents taux de sauts entre les autres états nous sont fournis dans un fichier .dat :

Listing 1 – taux_ass_hospit.dat

```
"AGE" "MU_AH" "MU_HA" "MU_AG" "MU_GA"
"1" 20 0.000109 0.2031 1.2e-05 0.138789
"2" 21 0.00011 0.1722 2.6e-05 0.138789
"3" 22 0.000111 0.162 5e-05 0.138789
"4" 23 0.000113 0.1623 8.6e-05 0.138789
...
"44" 63 0.000404 0.1187 0 0
"45" 64 0.00041 0.1175 0 0
"46" 65 0.000416 0.1180 0 0
```

Voici la description de chacun des fichiers :

0_function_markov.r contient l'ensemble des fonctions reçues avec le sujet du projet (corrigé de quelques petites coquilles)

1_mise_en_place.r contient l'ensemble des paramètres "standards" ainsi que le formatage de la matrice des taux de sauts et des fonctions simples permettant de mener à la **prime_pure_unique**

2_function_estimation.r contient 3 fonctions d'estimation (prime pure et de quantile)

3_reponses_aux_questions.r les réponses détaillées des questions 1 à 9

4_reponses_aux_questions.r les réponses détaillées des questions 10 à 11

I Mise en place

Cette partie concerne l'initialisation des différentes variables telles que :
- le taux technique $i = 4\%$
- le risque d'erreur $\alpha = 5\%$
- le nombre de jours dans l'année 365.25

Enfin, on a rassemblé les données brutes concernant les taux de sauts dans une matrice de dimension 3, $matrice\_taux[a, b, t]$ telle que :
- $t \in [[1; 46]]$ représente l'âge ($t = 1 \iff age = 20$)
- $a \in [[1; 3]]$ représente la ligne de la matrice de taux de sauts numéro t
- $b \in [[1; 3]]$ représente la colonne de la matrice de taux de sauts numéro t

On pourra trouver à la fin de cette partie du programme quelques fonctions permettant de faciliter la compréhension du programme et d'alléger le code. Celles-ci sont entièrement détaillées dans le code via l'en-tête de chaque fonction.

II Fonction d'estimation

II.1 Les arguments

Après avoir mis en place l'ensemble des outils et paramètres nécessaires, j'ai donc implémenté la fonction permettant l'estimation de la prime pure d'un contrat, celle-ci se nomme $prime\_pure\_unique$ et prend les arguments suivants :

age $\in [[20;65]]$ est l'âge (entier) de l'assuré à la souscription du contrat (valeur par défaut : 20)

T $\in [[1;66-age]]$ est la durée de couverture du contrat (valeur par défaut : NA)
  si la valeur de T est renseignée à NA alors la fonction va utiliser la durée de souscription maximale et renvoyer une alerte

i est le taux technique (valeur par défaut : 4%)

$indemnisations$ est le vecteur d'indemnisation contenant pour chaque numéro d'état le montant de l'indemnité annuelle (valeur par défaut : $c(0,171,164)*nombre\_de\_jours\_dans\_lannee$)

$matrice\_des\_taux$ est la matrice contenant les valeurs des taux de sauts pour chaque âge (valeur par défaut : $matrice\_taux$)

$func\_init$ est la fonction permettant de définir dans quel état l'assuré se trouve à la souscription (valeur par défaut : $function(t)return(1)$, c'est l'état valide)

$delai\_de\_carence$ la part d'année qui correspond à la durée que l'Assureur ne prendra pas en charge à chaque sinistre (valeur par défaut : 0)

$franchise$ est le montant annuel que l'Assureur va demander à l'assuré en cas de sinistre sur l'année calendaire (valeur par défaut : 0)

$detail$ est un paramètre permettant le choix de renvoi ou non du détail des calculs de flux (valeur par défaut : $FALSE$)

II.2 Exemple

Pour une assurée ayant 28 ans et souhaitant souscrire jusqu'à ses 40 ans sans franchise mais avec 1 jour de délai de carence (nous souhaitons afficher le détail du résultat), il suffira d'appliquer la fonction suivante :

Listing II.1 – Exemple d'utilisation de la fonction de prime pure unique

```
> prime_pure_unique(age = 28,
                  T = 12,
                  delai_de_carence = 1 / nombre_de_jours_dans_lannee,
                  detail = TRUE)
$age
[1] 28

$T
[1] 12

$pp_unique_avec_carence_et_franchise
[1] 3149.191

$pp_unique
[1] 3459.288

$info_sinistres
     date duree_sans_carence duree_dindemnisation type_indemnite_
        annuelle
1 1.573395   0.0409074072      0.03816956          62457.75
2 3.647076   0.0202696112      0.01753176          59901.00
3 5.802167   0.0001448308      0.00000000          62457.75

$flux_annuel
  date cout_sinistre_en_zero part_assureur_en_zero part_carence_en_zero
1    1          2400.157659        2239.3996       160.758010
2    3          1051.925534         909.7913       142.134221
3    5             7.204695           0.0000         7.204695
  part_franchise_en_zero part_franchise_en_zero_maxi
1                      0                         0
2                      0                         0
3                      0                         0

$chaine_simulee
$chaine_simulee$dates
[1] 0.000000 1.573395 1.614302 3.647076 3.667345 5.802167 5.802311
[8] 12.000000

$chaine_simulee$etats
[1] 1 2 1 3 1 2 1 1
```

III Analyse

Pour suivre l'évolution du code au fil de la lecture de ce rapport, jusqu'à la question 9 il suffit simplement de faire :

```
source("3_REPONSES_AUX_QUESTIONS.R")
```

puis à partir de la question 10 :

```
source("4_REPONSES_AUX_QUESTIONS.R")
```

Pour passer à la question suivante étape après étape, il suffira de cliquer sur la touche entrée à chaque étape. Pour désactiver cette option, il suffit de modifier le code de mise en place (ligne 9) :

```
# Step by step
sbs <- FALSE
```

III.1 Engagements de l'Assureur sur 1 an ($T = 1$)

On rappelle que la fonction **prime_pure_unique** permet de simuler pour une assurée en fonction des différents paramètres décrits plus haut, la valeur des engagements de l'Assureur ($primes\_pure\_unique$) sur l'ensemble de la période de couverture du contrat.

$$primes\_pure\_unique = \int_{t=0}^{T} e^{-rt} * (i_h * X_{t,h} + i_g * X_{t,g})\, dt$$

avec :

- $r = log(1 + i)$ le taux d'actualisation continu
- $i_h = 171$ euros par jour en cas d'hospitalisation (hors grossesse)
- $i_g = 164$ euros par jour en cas d'hospitalisation pour grossesse
- $X_{t,h} = 1$ si l'individu est hospitalisé (hors grossesse) en t, 0 sinon
- $X_{t,g} = 1$ si l'individu est hospitalisé pour grossesse en t, 0 sinon

Pour obtenir une estimation des engagements moyens de l'Assureur pour chaque âge (A), on va donc effectuer un grand nombre de simulations ($n = 100000$) de prime pure unique pour un âge ($primes\_pure\_unique_{A,n}$) puis en faire la moyenne (PP_A).

$$PP_A = \sum_{k=1}^{n} primes\_pure\_unique_{A,n}$$

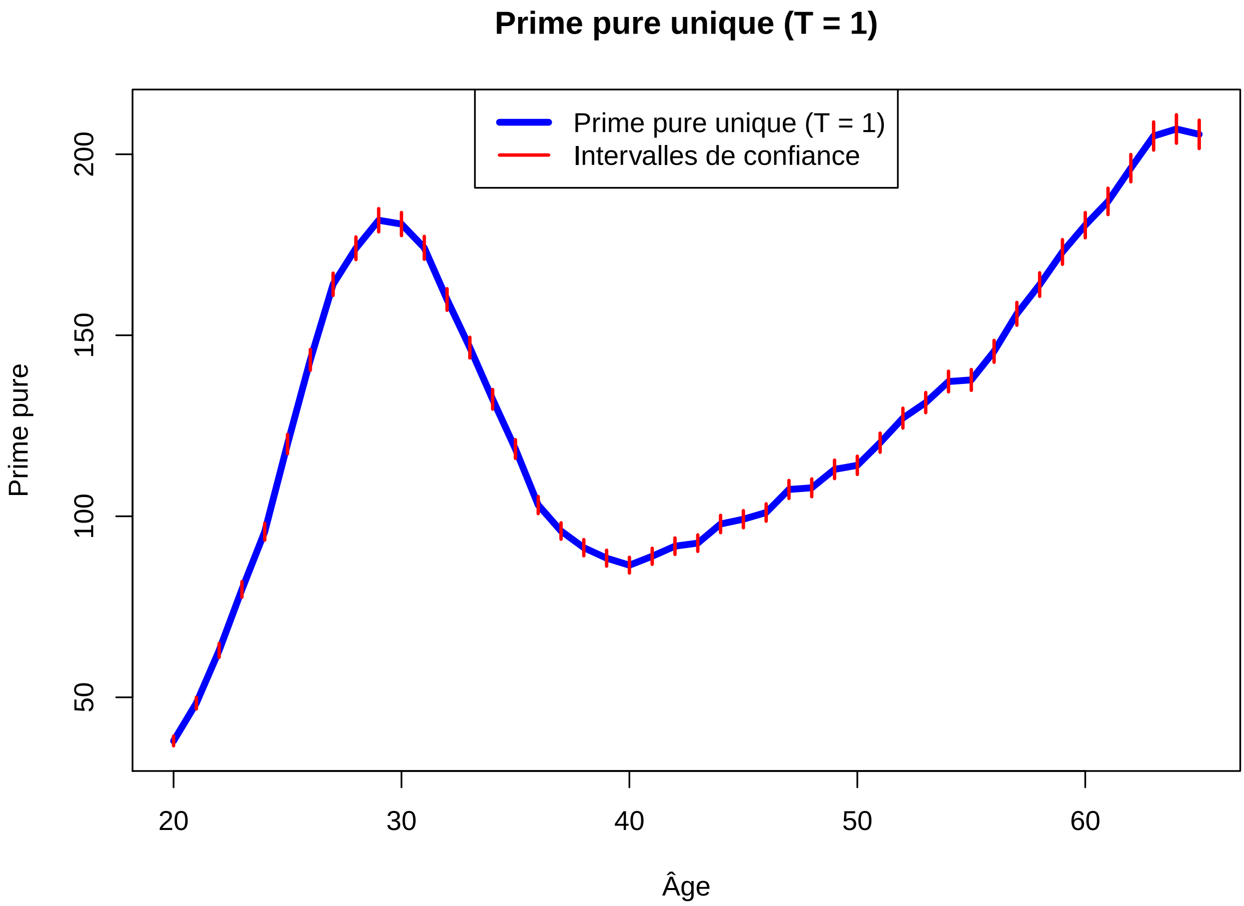
D'apès le Théorème Central Limite, on a l'intervalle de confiance non paramétrique suivant :

$$IC_\alpha = \left[IE(X) - t_\alpha * \sqrt{\frac{Var(X)}{n}} ; IE(X) + t_\alpha * \sqrt{\frac{Var(X)}{n}} \right]$$

t_α le quantile d'ordre α de la loi $N(0; 1)$

III.2 Prime pure unique par an $(T = 1)$

D'après l'équité actuarielle, les engagements de l'Assureur doivent être égaux aux engagements de l'assuré en date 0 (souscription du contrat), ce qui nous donne alors le graphique des primes pures uniques suivant :



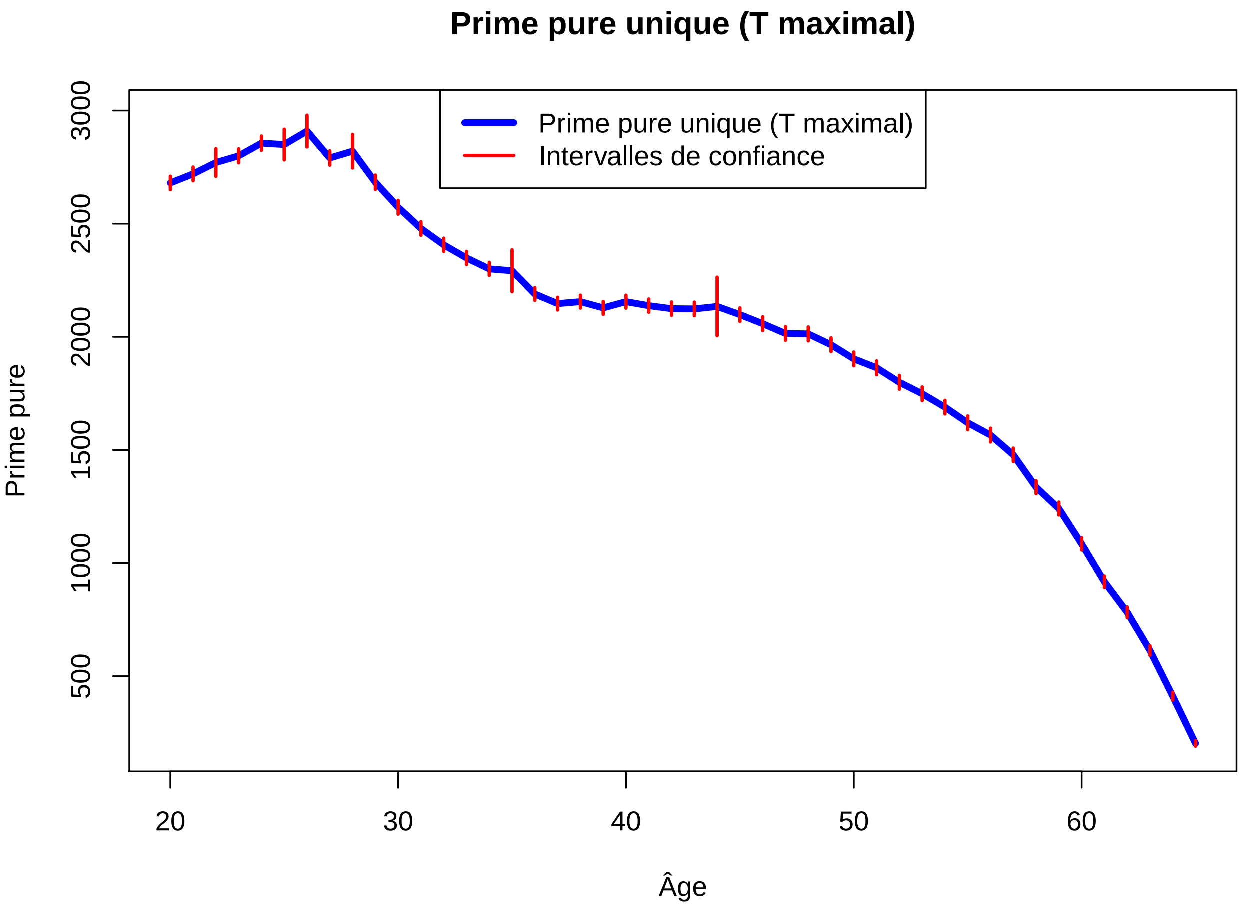
Tout d'abord, on remarque un pic à l'âge 30 ans, après analyse des matrices de taux on peut affirmer qu'il est dû à l'augmentation de la probabilité de tomber enceinte. Le creux à l'âge de 40 ans représente le "croisement" entre la décroissance de la probabilité d'avoir un enfant et la croissance du risque d'aller à l'hôpital pour toute autre maladie.

Il faut tout de même nuancer les choses, car dans notre modèle nous avons une indemnisation journalière supérieure dans le cas d'une maladie, ce qui met plus de "poids" (pour cette courbe) sur le risque de maladie.

III.3 Prime pure unique pour un contrat jusqu'à l'âge maximum

Dans ce cas, on va simplement choisir une valeur de T correspondant pour chaque âge (A) à une durée de couverture jusqu'à la fin de l'année des 65 ans :

$$\boxed{T = 66 - A}$$



Cette courbe est très nettement décroissante ce qui s'explique par le fait que l'ensemble du risque d'un âge est contenu dans les âges inférieurs à celui-ci. Exemple : le risque sur $[40; 66[$ est contenu dans celui sur $[30; 66[$, la prime pure brute d'actualisation d'un assuré de 30 ans sera donc en moyenne supérieure à celle d'un assuré de 40 ans.

Les différentes périodes de croissances de cette courbe sont donc simplement dues au facteur d'actualisation qui ne permet pas de faire une simple somme.

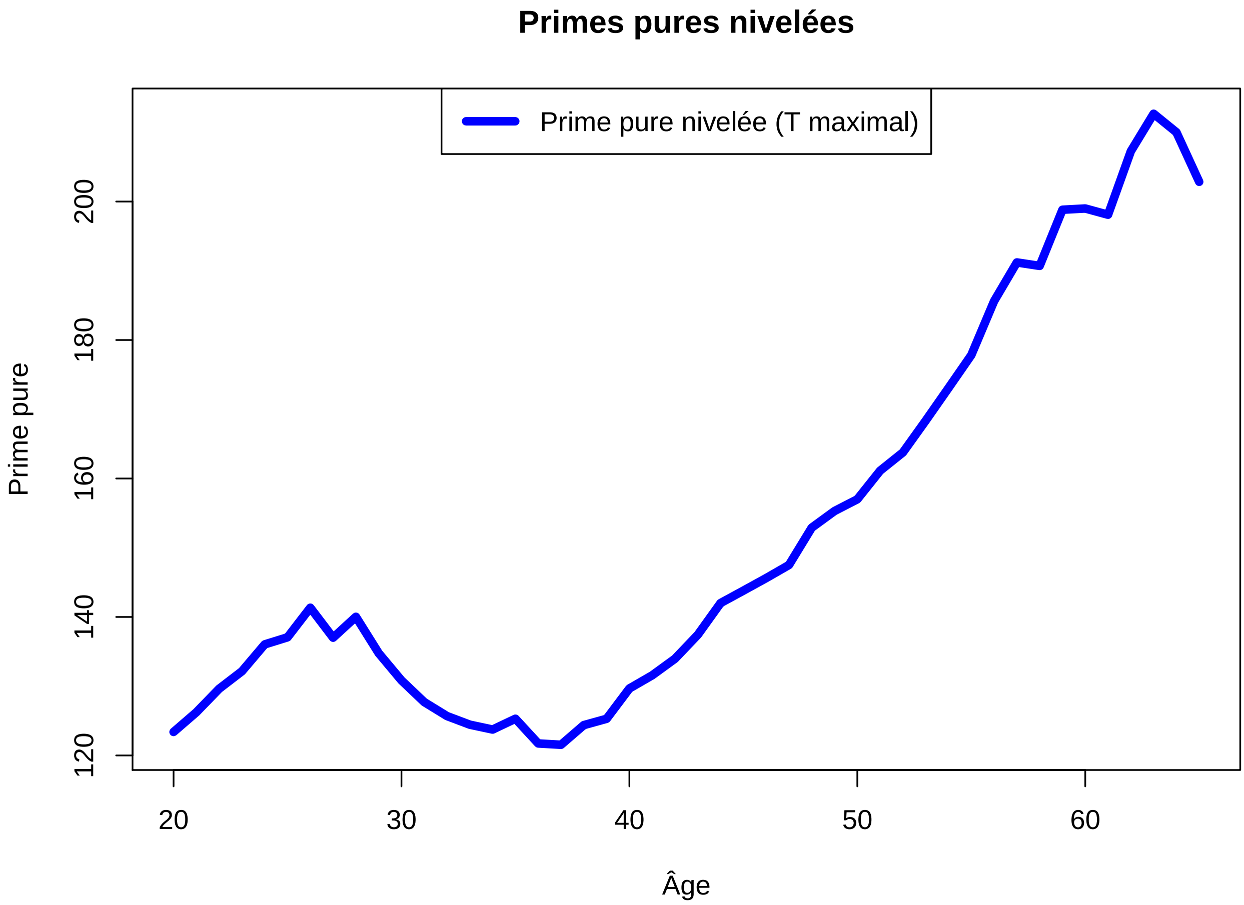
III.4 Prime pure nivelée pour un contrat jusqu'à l'âge maximum

La question précédente a permis de calculer la valeur des engagements de l'Assureur en date 0. Pour obtenir les primes pures nivelées, on va maintenant diviser cette valeur par le nombre de périodes équivalentes pendant lesquelles l'assuré va payer ses primes.

Pour une assurée d'âge A et un contrat d'une durée de $T = 66 - A$, le nombre d'annuités équivalentes pour un taux technique i est donné par :

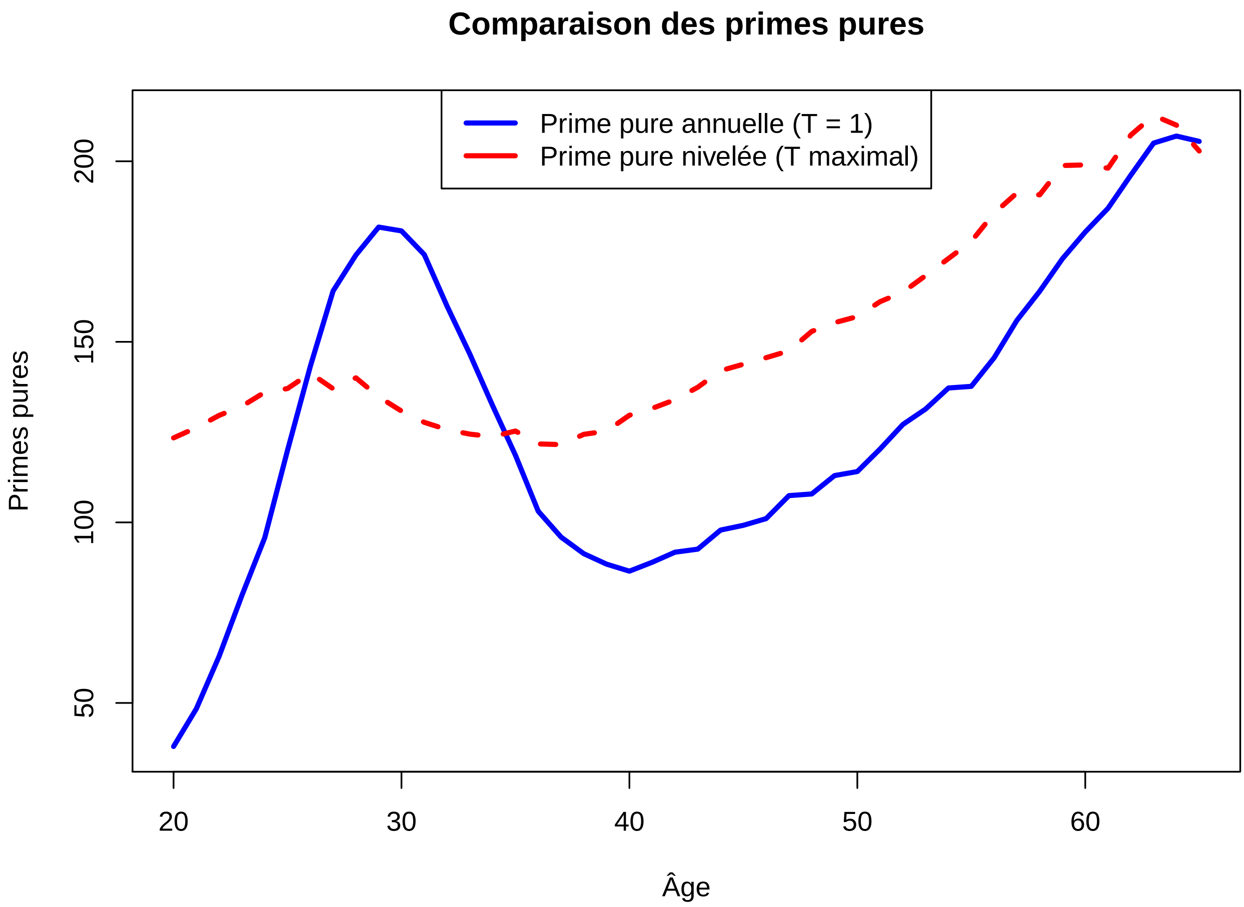
$$\ddot{a}_{\overline{n}|} = \sum_{k=0}^{T-1} \frac{1}{(1+i)^k} = \frac{1}{(1+i)^0} * \frac{1-\left(\frac{1}{1+i}\right)^T}{1-\frac{1}{1+i}} = \frac{1+i}{1+i} * \frac{1-(1+i)^{-T}}{1-\frac{1}{1+i}}$$

$$\boxed{\ddot{a}_{\overline{n}|} = \frac{1+i}{i} * \left(1-(1+i)^{-T}\right)}$$



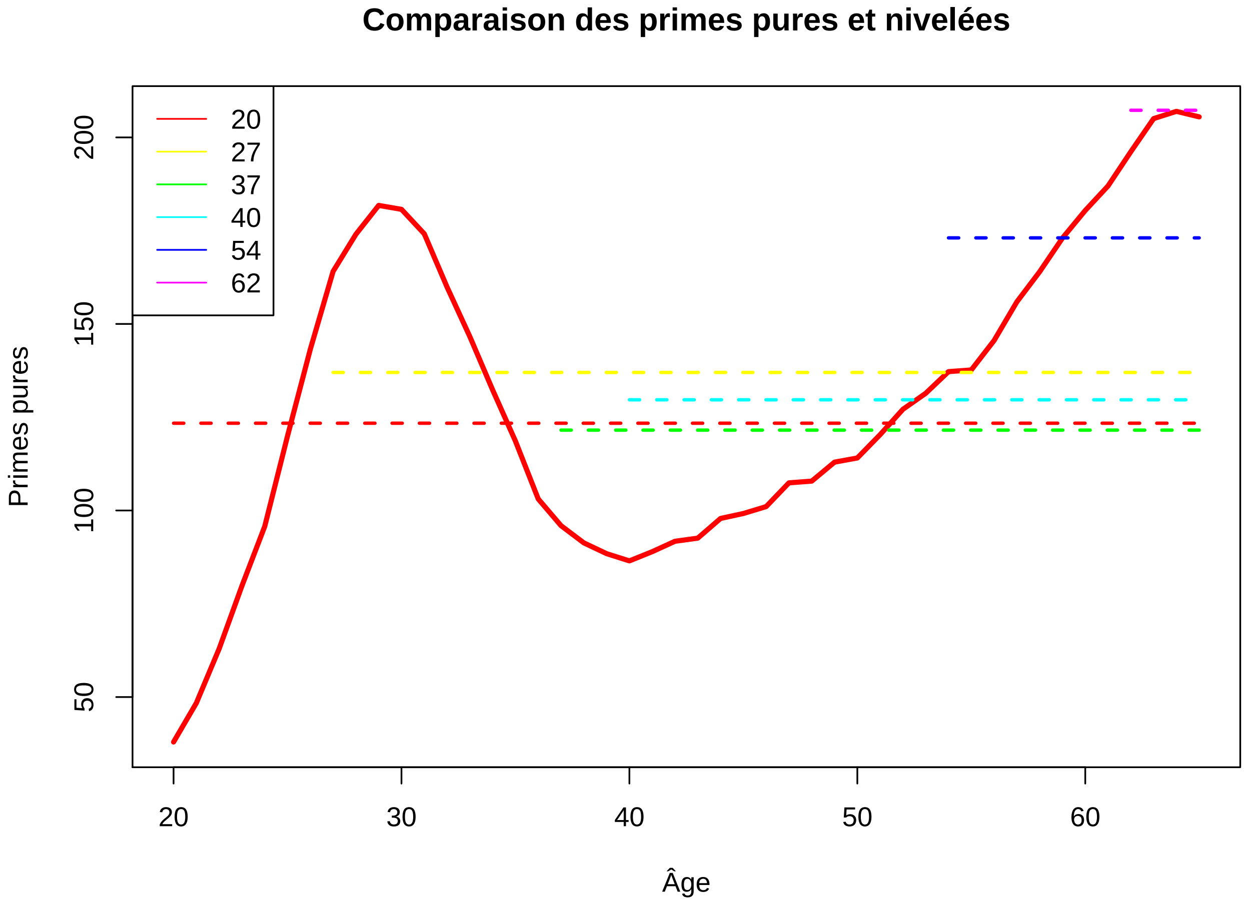
III.5 Comparaison des primes pures

Tout d'abord, voici les deux courbes de primes pures sur un même graphique :

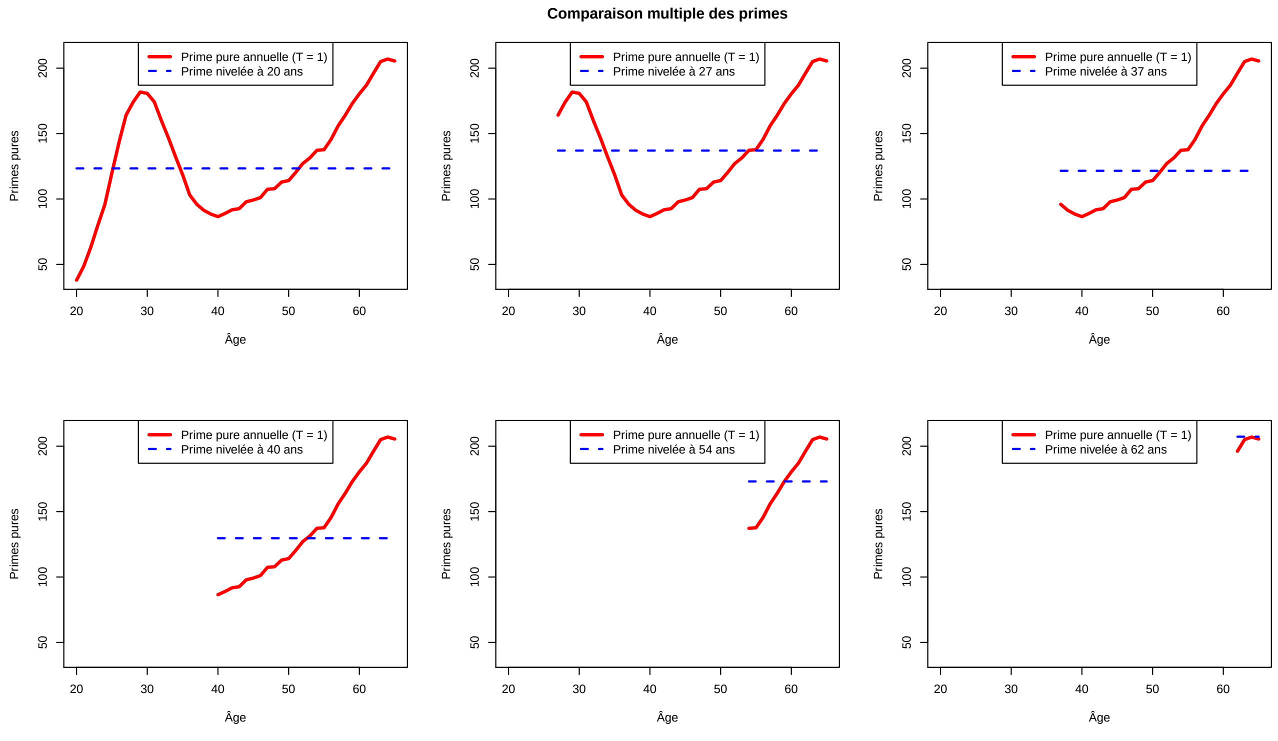


On remarque que l'allure est similaire avec un léger décalage à gauche et que très souvent les primes pures nivelées sont supérieures aux primes annuelles, ce à quoi on pouvait s'attendre car la prime nivelée représente l'ensemble du risque jusqu'à l'âge de 66 ans et la prime annuelle seulement sur une durée de 1 an.

Dans le but de mieux comparer les primes pures annuelles et nivelées, j'ai décidé de présenter les courbes pour 6 des âges possibles :



On remarque un minimum de prime nivelée global à 121 atteint en 20 et 37 ans, un maximum local de 139 à 27 ans, puis une croissance entre 37 et 65 ans. Ci-dessous se trouvent les 6 graphiques permettant de comparer les courbes des primes des 6 âges précédents.

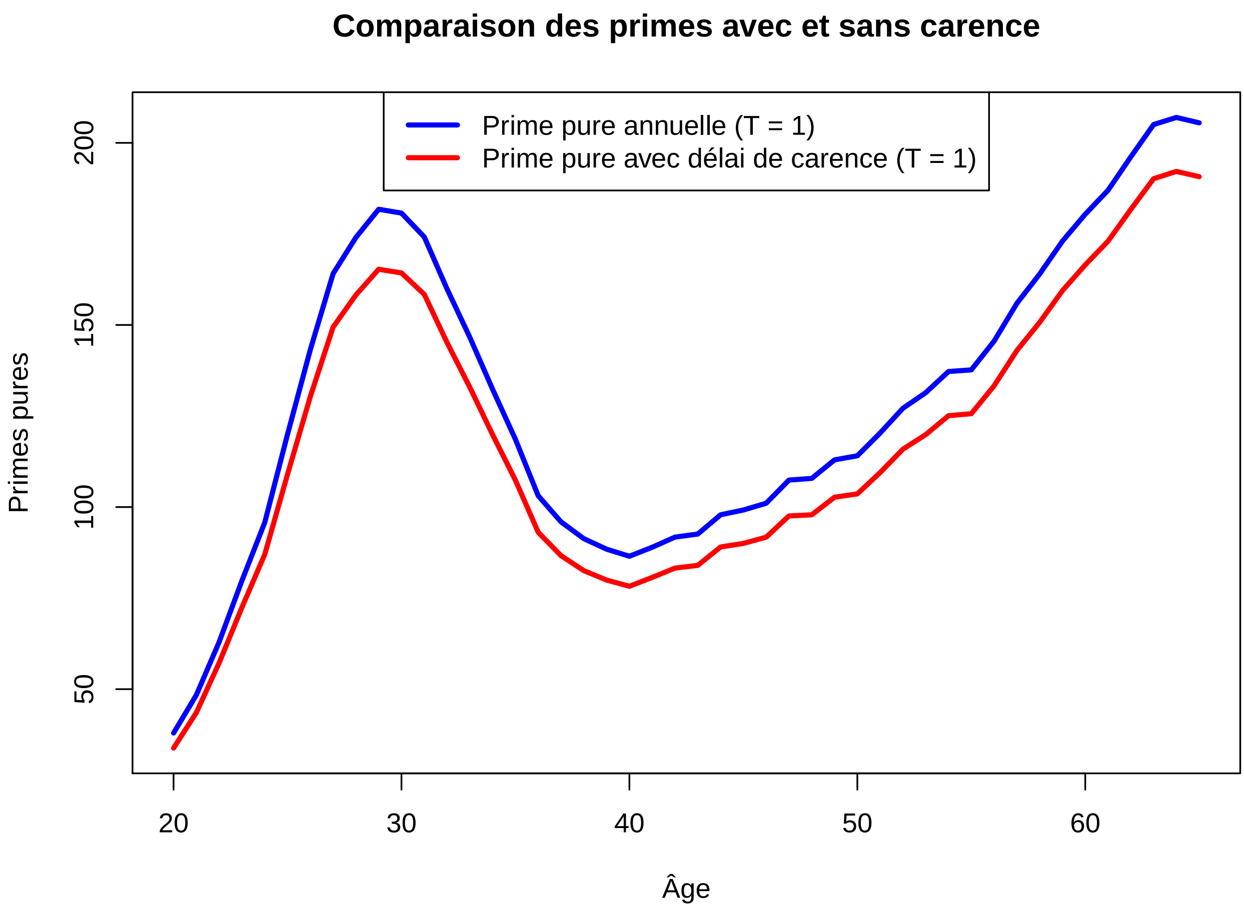


III.6 Avec délai de carence

La durée d'hospitalisation n'est donc plus comptée dès l'entrée en clinique mais à partir du délai de carence (d), c'est cette réduction de durée d'indemnisation qui abaisse la courbe des primes pures. Pour un sinistre ayant débuté en t_1 et finissant en t_2 :

$$cout\_de\_sinistre = \int_{t=t_1+d}^{t_2} e^{-rt} * (i_h * X_{t,h} + i_g * X_{t,g}) \, dt$$

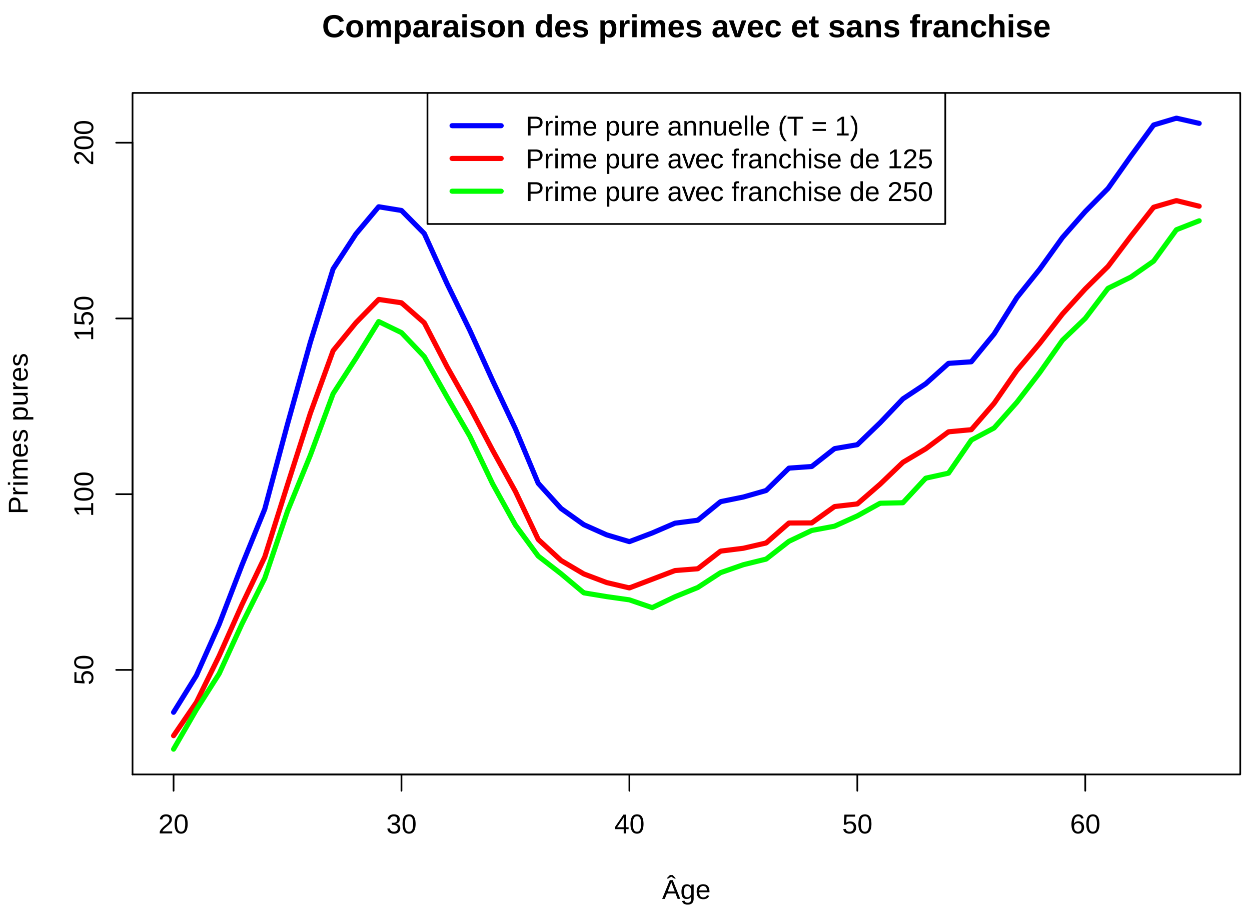
Voici la courbe que l'on obtient en ajoutant 1 jour de délai de carence pour chaque sinistre :



III.7 Avec franchise

La valeur de l'indemnité annuelle va être simplement abaissée de la valeur de la franchise ($franchise$) comptée en début d'année. Pour un sinistre entre t_1 et t_2 (dates calculées depuis le début de l'année) :

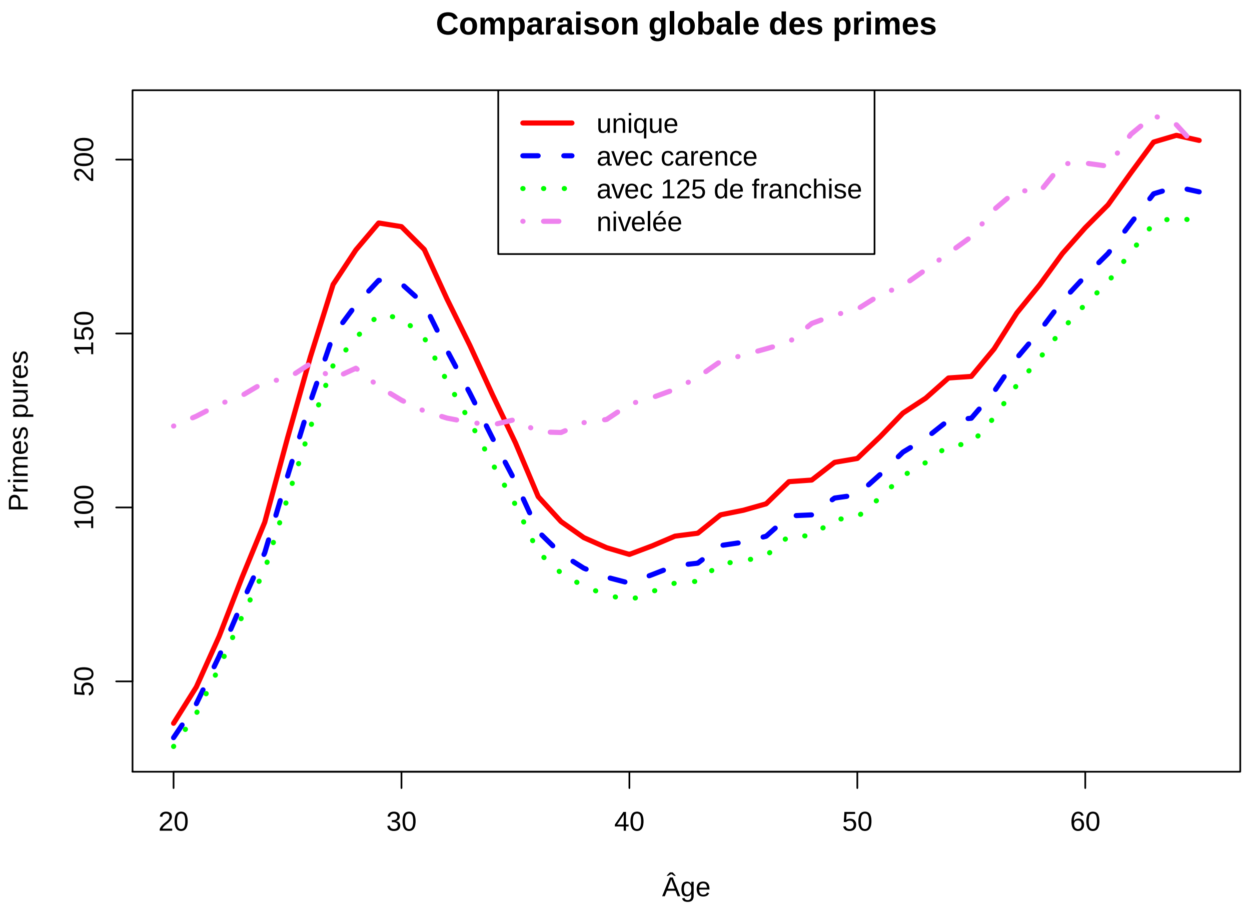
$$cout\_de\_sinistre = max(\int_{t=t_1}^{t_2} e^{-rt} * (i_h * X_{t,h} + i_g * X_{t,g})\ dt - franchise; 0)$$



On remarque que le fait de doubler la franchise ne réduit pas proportionnellement la valeur de la prime pure. Ceci peut s'expliquer par la quantité importante de sinistres courts et donc de faible indemnisation. 125 permet donc de ne pas tenir compte des "petits" sinistres et 250 abaisse légèrement le niveau de prestation.

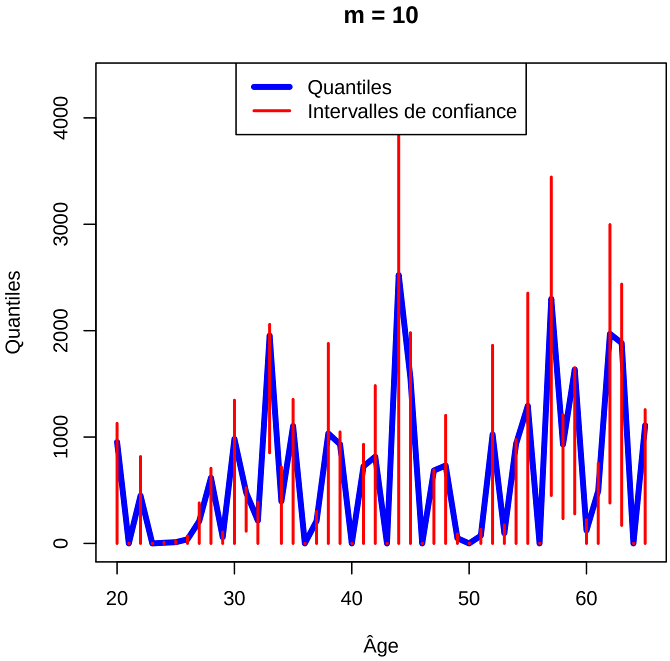
Point de comparaison

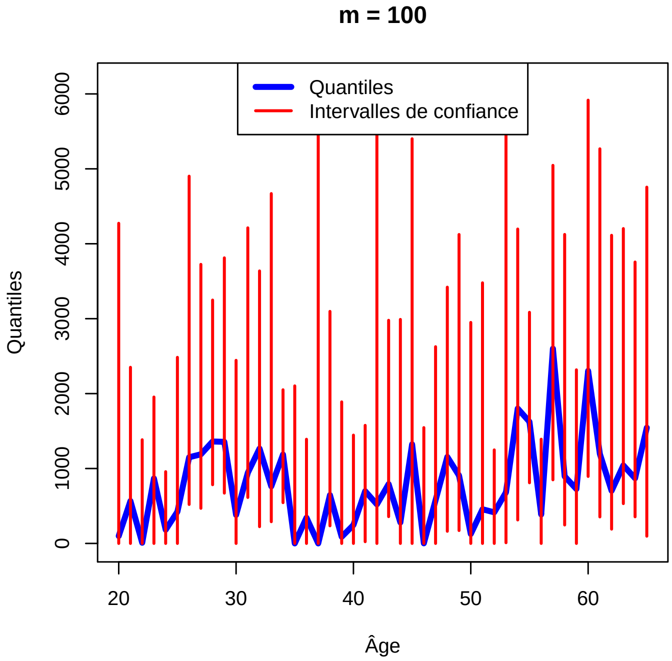
Voici donc le graphique de comparaison de l'ensemble des primes précédemment calculées :

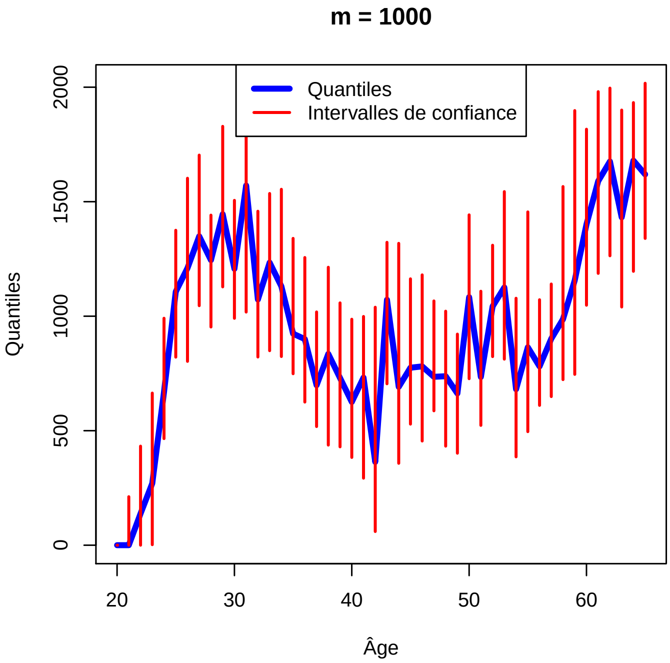


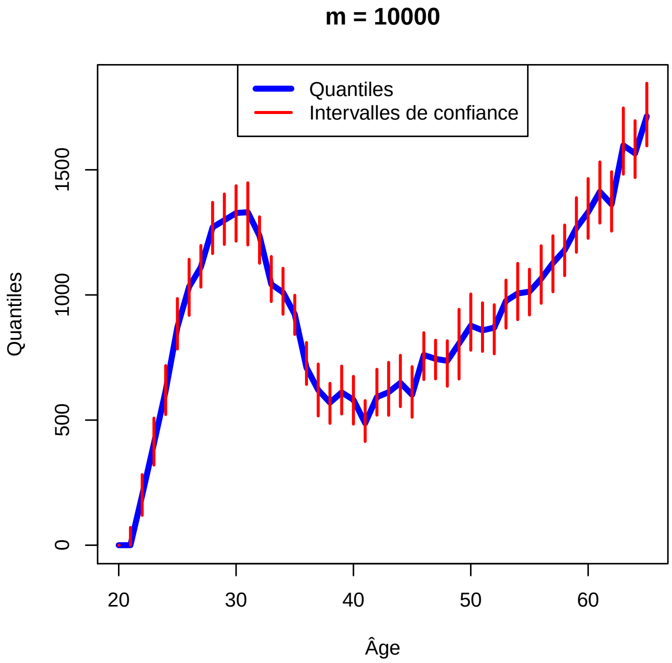
III.8 Estimation de quantile

On se place dans le cadre d'un contrat d'une durée d'un an et on estime le quantile d'ordre 95% par la méthode non paramétrique détaillée dans le cours.









On remarque que pour n = 10 et n = 100, les estimations sont peu intéressantes car très volatiles. On remarque que la tendance de la courbe est globalement la même que celle de l'ensemble des courbes précédentes. De plus, on constate graphiquement, au vu de la courbe n = 10000, que le coût par sinistre maximum au niveau de risque 5% est de l'ordre de 1800 euros sur une année.

III.9 Application de choc

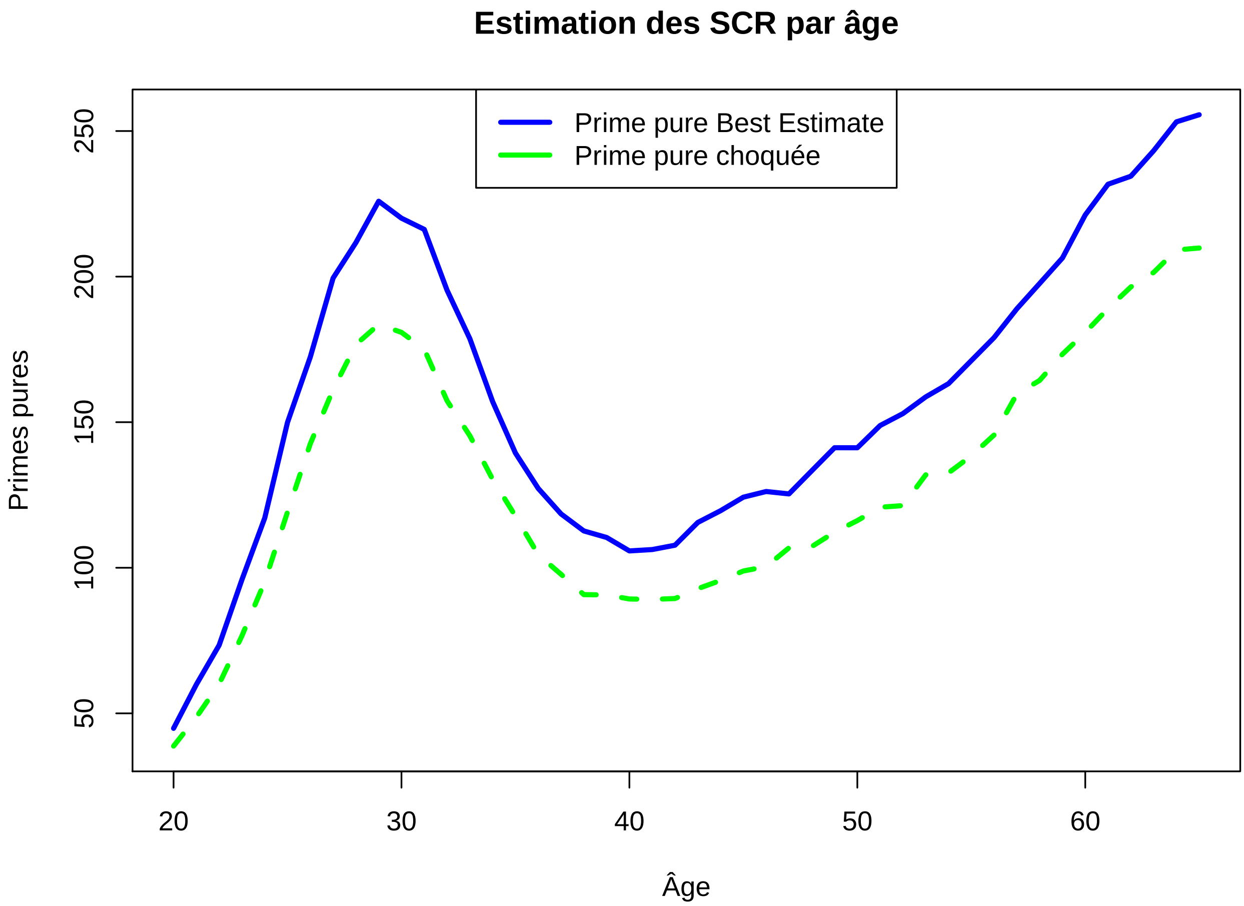
Tout d'abord, l'ensemble des paramètres non cités ci-dessous étant fixés par contrat, que nous supposeront sans clause de révision, les différents chocs sont applicables aux paramètres suivants :

- taux de sauts :
  - l'augmentation des taux de l'état hospitalisé (2 ou 3) vers actif (1) va baisser la prime pure
  - l'augmentation des taux de l'état actif vers hospitalisé va augmenter la prime pure
- le taux technique, son augmentation va faire diminuer la prime pure

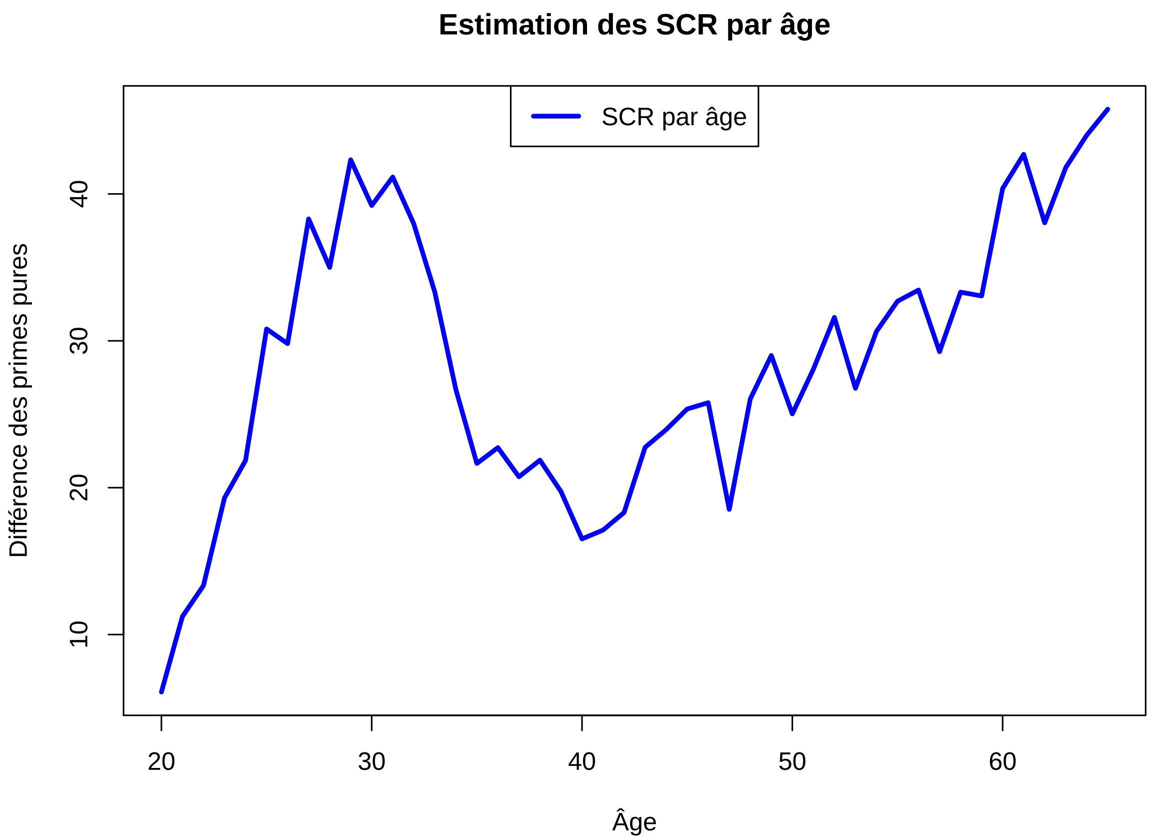
Pour appliquer ces chocs, il suffit simplement de modifier les valeurs des paramètres utilisés dans les fonctions, d'appliquer les calculs puis de constater les variations d'engagements.

Par exemple :
Supposons que les experts aient décidé d'appliquer un choc de +10% aux taux de passages de l'état actif à hospitalisé et de -10% à leur opposé permettant de passer des taux Best Estimate aux taux permettant l'évaluation des SCR.

On peut donc recalculer la matrice des taux puis réestimer les primes pures, ce qui nous donne le résultat suivant :



Et voici donc la différence de ces deux estimations, ce qui nous permet de connaitre la somme a provisionner par âge et par contrat pour faire face à nos engagements futurs en cas de choc :



III.10 Estimation des matrices de taux

Rappel : pour continuer dans le code :

```
source("4_REPONSES_AUX_QUESTIONS.R")
```

Pour estimer les taux de sauts, on va procéder de la manière suivante. Tout d'abord, on va simuler un grand nombre (n = 10000) de chaînes représentant chacun des individus. Pour chaque individu on va alors calculer par âge ($Ent(t)$) le nombre (N) de passage d'un état (a) à un autre (b) :

$$\boxed{N[a, b, Ent(t)] = \sum_{j=1}^{n} \sum_{t=0}^{T} n_{j,(a->b),t}}$$

- $n_{j,(a->b),t}$ vaut 1 si l'individu j est passé de l'état $a \in [[1;3]]$ à l'état $b \in [[1;3]]$ au temps t et 0 sinon
- $Ent(t)$ étant la partie entière de t

Ensuite, calculons le temps passé dans chacun des états (a) par l'ensemble des individus (j) en fonction de leur âge (k) :

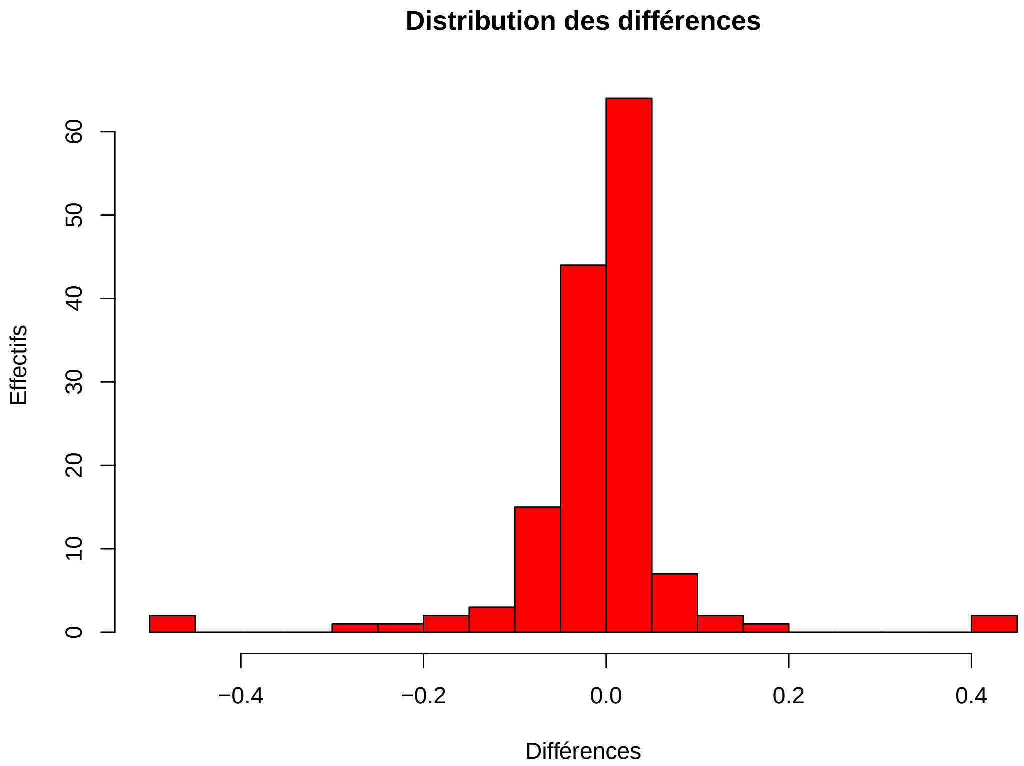
$$\boxed{T[k, a] = \sum_{j=1}^{n} T_{j,a,k}}$$

- $T_{j,a,k}$ est le temps que l'individu j a passé dans l'état a tout au long de l'année de ces k ans

Pour trouver le taux de saut à l'âge k du passage d'un état (a) à un autre (b), il suffit de diviser le nombre de passages par l'exposition au risque :

$$\boxed{\mu[a,b,k] = \frac{N[a,b,k]}{T[k,a]}}$$

Après calcul voici la distribution des différences entre les taux bruts et leurs estimations relatives :



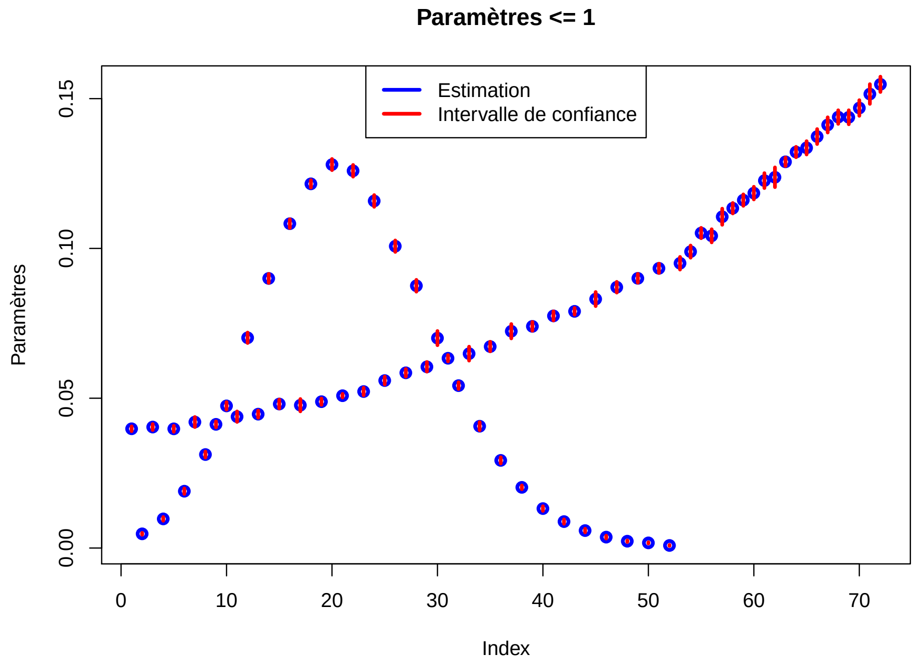

```
> t(t(summary((mu_estime - matrice_taux) / matrice_
   taux)))
       [,1]
Min.    -0.49760
1st Qu. -0.02838
Median   0.00344
Mean    -0.00637
3rd Qu.  0.02515
Max.     0.44060
NAs    270.00000
```

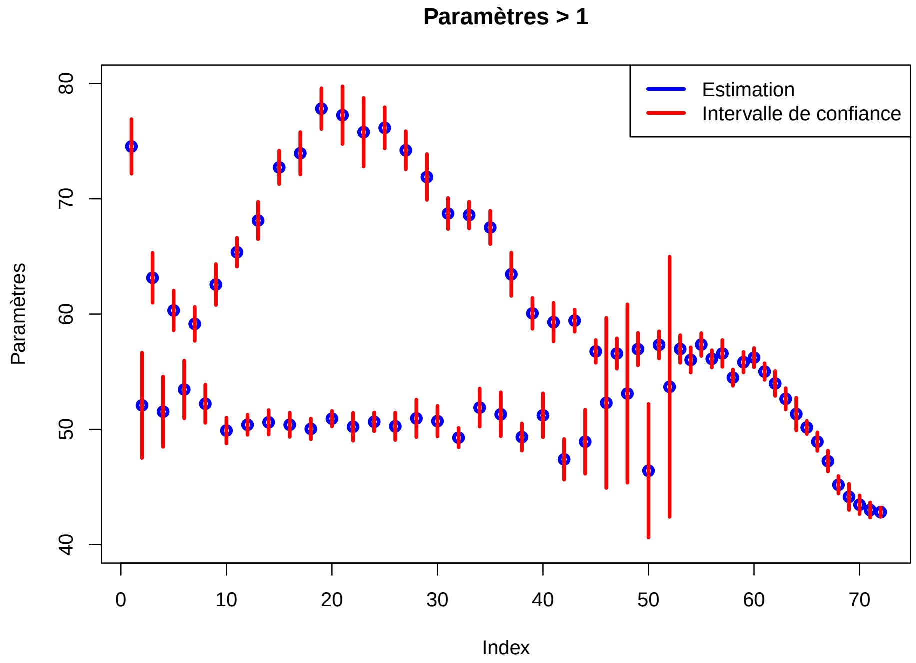
On remarque par la distribution des différences que l'estimation n'est pas totalement aberrante. Pour augmenter cette précision, il faudrait malgré tout augmenter le nombre de simulations, ce qui propose problème en pratique, car le nombre d'assurées en portefeuille n'est malheureusement pas aussi élevé que cela.

III.11 Incertitude sur les paramètres

Pour mesurer l'incertitude sur l'estimation des différents paramètres, on va répéter l'étape précédente 10 fois et calculer moyennes et variances pour chaque paramètre. Par application du Théorème Central Limite on peut obtenir un intervalle de confiance sur l'estimation des paramètres (cf section III.1).

Voici les graphiques contenant les estimations des paramètres non nuls ainsi que leurs intervalles de confiance respectifs :





III.12 Incertitude sur les primes

L'incertitude sur l'estimation des différents paramètres va directement influer sur la hauteur des engagements de l'Assureur. En effet, plus la volatilité de ces estimations est grande plus le portefeuille sera risqué, il faudra donc appliquer une marge de sécurité plus importante au risque de perdre des clients.

Malheureusement l'estimation par chaîne de Markov ne permet pas de trouver une relation fonctionnelle a priori entre les primes et les taux, ce qui ne permet pas de mesurer simplement l'impact de l'incertitude sur les taux appliqués aux primes. Pour étudier le phénomène, on pourra par exemple procéder comme dans la section III.9 en appliquant des chocs aux taux permettant d'évaluer les primes maximum et minimum.

III.13 Les facteurs d'incertitude

Tout d'abord, le premier paramètre d'intertitude est directement lié à la taille de l'échantillon car plus il est réduit, moins il est représentatif de la population. La seule solution est d'augmenter le nombre de personnes étudiées ce qui en pratique est très lié à la taille du portefeuille, souvent assez réduite.

Après l'estimation des taux d'après le point précédent et avec l'utilisation des chaînes de Markov, pour augmenter la précision du calcul des engagements Best Estimate ou de la Value at Risk, il faudra augmenter le nombre de simulations.

Concernant le taux technique, il peut être intéressant de réaliser une étude sur l'ensemble des valeurs possibles du taux technique. Cet ensemble pourra par exemple se baser dans un premier temps sur l'ensemble historique.

IV Conclusion

Cette étude a donc permis le calcul des engagements de l'Assureur selon de multiples cas, notamment avec délai de carence et/ou franchise. Le tout étant paramétrable via la fonction principale : **prime_pure_unique** .

Un prolongement de cette étude pourrait être l'analyse détaillée des taux de passages, car on peut remarquer que les primes sont très fortement influencées par ces taux. Il faudrait nottement évaluer les différents niveaux de chocs à appliquer par exemple dans le cadre de Solvabilité II.