**EXAMEN FINAL, ACT2040, HIVER 2013**

Les calculatrices[1] sont autorisées, mais pas les téléphones 'intelligents' (qui devront être rangés pendant toute la durée de l'épreuve). Tous les documents sont interdits. Comme rappelé dans le plan de cours, la triche est interdite.

Dans les feuilles qui suivent, il y a 25 questions,

- 19 questions sur le provisionnement pour sinistres à payer (total de 22 points)
- 6 questions portant sur des aspects de tarification (mais comptant pour 15 points)

**Précisions sur la notation** Pour chaque question, une courte réponse doit être apportée, détaillée si demandé. Vous gagnez les points si la réponse est claire, et correcte. Si aucun détail n'est demandé, seule la bonne réponse apportera le point. Votre note finale est le total des points (sur 37).

**Précisions sur l'épreuve** Je ne réponds à aucune question durant l'épreuve. En cas de doute, ou de suspicion d'erreur, merci de m'en faire part dans le cahier d'examen.

## 1. Méthodes de provisionnement

On a observé le triangle de paiements suivant (la colonne de gauche indiquant l'année)

Triangle 1

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 2007 | [redacted] | 9668 | 10563 | 10771 | 10978 | 11040 |
| 2008 | 6346 | 9593 | 10316 | 10468 | 10536 | |
| 2009 | 6269 | 9245 | 10092 | 10355 | | |
| 2010 | 5863 | 8546 | 9268 | | | |
| 2011 | 5778 | 8524 | | | | |
| 2012 | 6184 | $\star$ | $*$ | | | |

[1] BA-35, BA II Plus, TI-30X, TI-30Xa, TI-30XIIS et TI-30XIIB (cf plan de cours).

**Question 1.** Quel montant a été payé en 2012 pour les sinistres survenus en 2010 ? [*Le calcul sera explicité dans la feuille.*]

Pour les sinistres survenus en 2010, fin 2010, on avait payé 5863; fin 2011 on avait payé 8546; et fin 2012, 9268. Autrement dit, pendant l'année 2012, 9268-8546 ont été payé, soit 722

**Question 2.** Le premier montant, en haut à gauche, a été effacé. En utilisant la méthode Chain Ladder, quel montant serait-il réaliste d'avoir ? [*Le calcul sera explicité dans la feuille.*]

L'idée de la méthode Chain Ladder est que $C_{i,j+1} = \lambda_{j+1} \cdot C_{i,j}$ (ou avec $\lambda_j$, suivant les notations , mais peu importe, ça ne change pas les calculs). Ou dit autrement, $C_{i,j} = \lambda_{j+1}^{-1} \cdot C_{i,j+1}$. Donc si on peut estimer le coefficient de transition entre la première colonne (0) et la seconde (1), on pourra extrapoler la valeur manquante. On peut utiliser les données dont on dispose, sur les lignes correspondant aux années 2008 à 2011,

$$\lambda_1 = \frac{9593 + 9245 + 8546 + 8524}{6346 + 6269 + 5863 + 5778} = \frac{35908}{24256} = 1.480376$$

On peut alors extrapoler en considérant

$$\lambda_1^{-1} \cdot C_{2007,1} = \frac{9668}{1.480376} = 6530.773$$

Il serait réaliste d'avoir un montant de l'ordre de 6531.

**Question 3.** Quel serait le montant qui devrait être prédit, par la méthode Chain Ladder pour 2012, en $\star$ ? [*Le calcul sera explicité dans la feuille.*]

On nous demande d'extrapoler maintenant $\star$. Plusieurs se sont demandé si on devait utiliser la valeur extrapolée dans la question précédante. Dans un premier temps, on va tenir compte du fait que la valeur est ici inconnue. Dans ce cas, on utilise la valeur calculée à la question précédante pour le facteur de transition $\lambda_1$. En $\star$, on devrait prédire, par la méthode Chain Ladder

$$\lambda_1 \cdot C_{2012,0} = 1.480376 \cdot 6184 = 9154.645$$

Que se passerait-il si on utilisait la valeur extrapolée auparavant. Rien.... Car l'estimateur du facteur de transition est une moyenne pondérée de facteur de transition par année de survenance. Si on rajoute la moyenne (avec des poids proportionnels) sur les dernières lignes, on aura *exactement* le même facteur de transition. Pour les sceptiques, faisons le calcul:

$$\lambda_1 = \frac{9593 + 9245 + 8546 + 8524 + 9668}{6346 + 6269 + 5863 + 5778 + 6530.773} = 1.480376$$

Ce qui n'influencera pas la prédiction...

**Question 4.** Quel serait le montant qui devrait être prédit, par la méthode Chain Ladder pour 2012, en $*$ ? [*Le calcul sera explicité dans la feuille.*]

Pour $*$, il va falloir estimer $\lambda_2$. Ici, pas d'hésitation,

$$\lambda_2 = \frac{10563 + 10316 + 10092 + 9268}{9668 + 9592 + 9245 + 8546} = 1.086014$$

On en déduit alors (en utilisant la question précédante) question

$$C_{2012,2} = C_{2012,0} \cdot \lambda_1 \cdot \lambda_2 = 6184 \cdot 1.480376 \cdot 1.086014 = 9942.07.$$

**Question 5.** En supposant la première année close, quel serait le montant de provision que l'on devrait constituer pour l'année de survenance 2009, par la méthode Chain Ladder ? [*Le calcul sera explicité dans la feuille.*]

Pour 2009, il nous manque deux années de développement avant que l'année ne soit close. Pour les deux prochaines années, on devrait payer (on utilise, comme toujours, une simple règle de trois)

$$C_{2009,5} = C_{2009,4} \cdot \lambda_5 = C_{2009,3} \cdot \lambda_4 \cdot \lambda_5$$

soit

$$C_{2009,5} = C_{2009,3} \cdot \frac{C_{2007,4} + C_{2008,4}}{C_{2007,3} + C_{2008,3}} \cdot \frac{C_{2007,5}}{C_{2007,4}}$$

On remplace alors par les chiffres, pour l'application numérique,

$$C_{2009,5} = 10355 \cdot \frac{10536 + 10978}{10771 + 10468} \cdot \frac{11040}{10978} = 10548.31$$

Mais ce n'est pas fini, car ce montant n'est que la charge totale de sinistres que l'on pense payer pour les sinistres survenus en 2009. Comme on a payé, pour l'instant, 10355, le montant de provision estimé doit être le montant que l'on pense encore devoir payer dans les deux années à venir. Soit ici

$$R_{2009} = C_{2009,5} - C_{2009,3} = 10548.31 - 10355 = 193.31$$

**Question 6.** En réalité, l'année 2007 n'est pas close, et selon des dires d'experts, il convient de constituer une provision de 108, pour cette année de survenance. En tenant compte de cette information, quel serait votre nouvel estimateur pour le montant de provisions pour l'année de survenance 2009, par la méthode Chain Ladder ? [*Le calcul sera explicité dans la feuille.*]

Si la première année n'est pas close, et que 108 seront payées dans les années à venir. Bref, il faut utiliser un *tail factor* (pour reprendre les notations du cours) $\lambda_\infty$ tel que

$$\lambda_\infty = \frac{C_{2007,\infty}}{C_{2007,5}} = \frac{C_{2007,5} + R_{2007}}{C_{2007,5}} = \frac{11040 + 108}{11040} = 1.009783$$

Aussi, pour 2009, pour extrapoler la charge finale, il faut prendre en compte ce facteur dans le calcul de la question précédante,

$$C_{2009,\infty} = C_{2009,5} \cdot \lambda_\infty = 10548.31 \cdot 1.009783 = 10651.51$$

Aussi, le montant de provisions à constituer est

$$R_{2009} = C_{2009,\infty} - C_{2009,3} = 10651.51 - 10355 = 296.5082$$

(soit presque 50% de plus).

Fin 2013, on dispose de davantage d'information. En particulier, on a retrouvé le payement effectué en 2007 pour les sinistres survenus en 2007. Et une année supplémentaire de paiements a été observée

Triangle 2

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 2007 | 5946 | 9668 | 10563 | 10771 | 10978 | 11040 | 11106 |
| 2008 | 6346 | 9593 | 10316 | 10468 | 10536 | 10572 | |
| 2009 | 6269 | 9245 | 10092 | 10355 | 10507 | | |
| 2010 | 5863 | 8546 | 9268 | 9459 | | | |
| 2011 | 5778 | 8524 | 9178 | | | | |
| 2012 | 6184 | 9013 | $*$ | | | | |

**Question 7.** Si nous avions connu ce montant fin 2012, le montant qui aurait du être prédit, par la méthode Chain Ladder pour 2012, en $\star$ serait il plus grand ou plus petit que ce qui a été calculé à la question 3. ? [*Une rapide explication sera donnée dans la feuille.*]

Ici, la question revient à refaire le calcul de la question 3 en tenant compte de la nouvelle observation. Soit

$$\lambda_1 = \frac{9668{+}9593 + 9245 + 8546 + 8524}{5946{+}6346 + 6269 + 5863 + 5778} = \frac{45576}{30202} = 1.509039$$

Donc si le montant avait été connu l'an passé, on aurait prédit pour $\star$

$$6184 \cdot \lambda_1 = 9331.898$$

qui est plus grand que le montant prédit à la question 3 (9154.645 en l'occurrence). En fait, comme le montant observé (5946) est plus petit que celui calculé dans la question 2 en extrapolant (6531), on aurait pu conclure directement (si le montant est plus faible, comme il est au dénominateur, on devrait avoir un plus grand facteur de transition)

**Question 8.** Sur ce nouveau triangle, quel serait le montant qui devrait être prédit, par la méthode Chain Ladder pour 2012, en $*$ ? [*Le calcul sera explicité dans la feuille.*]

Pour 2012, on se relance dans les calculs

$$\lambda_2 = \frac{10563 + 10316 + 10092 + 9268{+}9178}{9668 + 9592 + 9245 + 8546 + 8524{+}8524} = \frac{49417}{45576} = 1.084277$$

On en déduit alors

$$C_{2012,2} = C_{2012,1} \cdot \lambda_2 = 9013 \cdot 1.084277 = 9772.589.$$

**Question 9.** Sans tenir compte du tail factor mentionné dans la question 6, et en utilisant les calculs de la question 5, avait-on constitué assez de provision fin 2012 pour payer les sinistres de 2009 en 2013 ? [*Le calcul sera explicité dans la feuille.*]

On a payé 10507-10355 en 2013 pour les sinistres survenus l'année 2009, soit 152. Or une provision de 193.31 avait été constitué fin 2012 (c'était la question 5). Donc oui, on avait assez provisionné (pour cette année au moins).

Dans les questions 10-12, on va utiliser le triangle suivant, constitué des $C_{i,j}$

Triangle 1b

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 2007 | 5946 | 9668 | 10563 | 10771 | 10978 | 11040 |
| 2008 | 6346 | 9593 | 10316 | 10468 | 10536 | |
| 2009 | 6269 | 9245 | 10092 | 10355 | | |
| 2010 | 5863 | 8546 | 9268 | | | |
| 2011 | 5778 | 8524 | | | | |
| 2012 | 6184 | | | | | |

ainsi que les données suivantes, correspondant aux primes acquises $\Pi_i$

Primes acquises

| | |
|---|---|
| 2007 | 15474 |
| 2008 | 14882 |
| 2009 | 14456 |
| 2010 | 14055 |
| 2011 | 14525 |
| 2012 | 15026 |

À partir du triangle de paiements, on calcule $S_{i,j} = C_{2007,j} + C_{2008,j} + \cdots + C_{i,j}$, donné dans le triangle suivant

Triangle 3

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 2007 | 5946 | 9668 | 10563 | 10771 | 10978 | 11040 |
| 2008 | 12292 | 19261 | 20879 | 21239 | 21514 | |
| 2009 | 18561 | 28506 | 30971 | 31594 | | |
| 2010 | 24424 | 37052 | 40239 | | | |
| 2011 | 30202 | 45576 | | | | |
| 2012 | 36386 | | | | | |

**Question 10.** À partir de ces données, calculer les facteurs de transition $\lambda_j$ par la méthode Chain Ladder, avec $C_{i,j} = \lambda_j \cdot C_{i,j-1}$ ? [*Aucun calcul ne sera détaillé sur la feuille, un tableau avec les différentes valeurs suffira*]

On a tout dans le tableau, car

$$\lambda_j = \frac{S_{2012-j,j}}{S_{2012-j,j-1}}$$

Par exemple pour $\lambda_1$, on utilise

$$\lambda_1 = \frac{45576}{30202} = 1.509039$$

qui correspond exactement à la valeur calculée dans la question 7, ce qui est rassurant car c'est le même triangle, mais aussi

$$\lambda_2 = \frac{40239}{37052} = 1.086014$$

puis

$$\lambda_3 = \frac{31594}{30971} = 1.020116$$

et encore

$$\lambda_4 = \frac{21514}{21239} = 1.012948$$

et pour finir

$$\lambda_5 = \frac{11040}{10978} = 1.005648$$

Bon, on a tout !

**Question 11.** Donnez le montant de provision nécesssaire, toutes années confondues, donné par la méthode Chain Ladder. [*Aucun calcul ne sera détaillé sur la feuille, seul le montant total de provision est demandé*]

Bon, ici il faut faire un peu de calcul, mais c'est facile,

$$R_{2008} = C_{2008,4} \cdot (\lambda_5 - 1) = 10536 \cdot (1.005648 - 1) = 59.5$$

$$R_{2009} = C_{2009,3} \cdot (\lambda_5 \cdot \lambda_4 - 1) = \cdots = 193.3$$

(je passe un peu calculs)

$$R_{2010} = C_{2010,2} \cdot (\lambda_5 \cdot \lambda_4 \cdot \lambda_3 - 1) = \cdots = 362.9$$

$$R_{2011} = C_{2011,1} \cdot (\lambda_5 \cdot \lambda_4 \cdot \lambda_3 \cdot \lambda_2 - 1) = \cdots = 1095.7$$

$$R_{2012} = C_{2012,0} \cdot (\lambda_5 \cdot \lambda_4 \cdot \lambda_3 \cdot \lambda_2 \cdot \lambda_1 - 1) = \cdots = 4347.4$$

Pour le montant total de provision, sommons (oui, c'est assez calculatoire comme exercice... c'est le principe de base de 95% des calculs actuarels),

$$R = 6058.89$$

**Question 12.** On suppose que, par année, le ratio sinistres sur primes doit ètre constant, de l'ordre de 72%. En utilisant la charge ultime qui donnerait un tel ratio, en déduire le montant qu'il faudrait provisonner, toutes années confondues. [*Aucun calcul ne sera détaillé sur la feuille, seul le montant total de provision est demandé*]

Ici, on a une méthode plus simple pour provisionner: on utilise le faire que $C_{i,\infty} = \Pi_i \cdot \text{S/P}$. Pour 2007, par exemple, on anticipe une charge total de $15474 \cdot 0.72 = 11141.28$. Comme on a payé 11040, il faut constituer une provision d'un montant 101.28. Pour 2008, on anticipe $14882 \cdot 0.72 = 10715.04$. Il faut donc une provision de l'ordre de $10715.04 - 10536 = 179.04$. C'est facile, non ? Calculatoire, mais facile... Bref, on continue. Pour 2009, $14456 \cdot 0.72 - 10355 = 53.32$. Pour 2010, etc. Le plus simple est de voir que la charge totale est la somme totale des primes, multplié par 72%, soit

$$(15474 + 14882 + \cdots + 15026) \cdot 0.72 = (88418) \cdot 0.72 = 63660.96$$

Ensuite, on calcule combien on a déja payé, en sommant la derniére diagonale,

$$(11040 + 10536 + \cdots + 6184) = 55907.00$$

Si on fait la différence, on a le montant total de provision, i.e. 7753.96 (que l'on comparera avec le montant obtenu avec la méthode Chain Ladder, 6058.89).

Dans la suite des questions 13-15, on va utiliser la sortie de régression suivante, basée sur le triangle 1b :

Régression 1

```
Call:
glm(formula = Y ~ i + j, family = quasipoisson, data = B)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.77688    0.05178 169.514  < 2e-16 ***
i2008       -0.04110    0.06871  -0.598   0.5630
i2009       -0.04556    0.06912  -0.659   0.5247
i2010       -0.13654    0.07128  -1.916   0.0844 .
i2011       -0.13771    0.07331  -1.878   0.0898 .
i2012       -0.04716    0.08237  -0.573   0.5796
j1          [redacted]   [redacted]  [redacted]  [redacted]
j2          -2.04177    0.09451 -21.605 1.01e-09 ***
j3          -3.41227    0.20468 -16.672 1.26e-08 ***
j4          -3.83292    0.30632 -12.513 1.97e-07 ***
j5          -4.64975    0.64187  -7.244 2.78e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for quasipoisson family taken to be 25.37786)
```

**Question 13** [×2]**.** Quelle est la valeur de l'estimateur masqué sur la sortie précédante, $\widehat{\beta}_1$ ? [*Le calcul sera explicité dans la feuille.*]

C'est une question de cours ça: quand on utilise les facteurs lignes et colonnes, dans un triangle complet, sur les incréments de payements (oui, j'avais oublié de préciser, mais un peu de relecture du cours et de bon sens permettait, je pense que comprendre que `Y` était l'incrḿent de payements), une régression de Poisson avec un lien logarithmique (ce qui est fait ici, car en terme de prédiction, `poisson` et `quasipoisson` coïncident) donne exactement les mêmes prédictions que la méthode Chain Ladder. Pas seulement le montant total de provision, mais aussi les prédictions de $C_{i,j}$ et des $Y_{i,j}$ dans la partie inférieure du triangle. Bref, il nous faut $\widehat{\beta}_1$ (tel que noté dans le cours). Le *seul* endroit où ce facteur intervient est sur les termes de la seconde colonne (1). En particulier

$$\widehat{Y}_{2012,1} = \exp[\widehat{\gamma} + \widehat{\alpha}_{2012} + \widehat{\beta}_1]$$

si $\gamma$ désigne la constante. Si on veut que la régression de Poisson coïncide avec la méthode Chain Ladder, il faut question $\widehat{Y}_{2012,1} = C_{2012,1} - C_{2012,0}$ où $C_{2012,1}$ est justement le montant prédit par la méthode Chain Ladder dans la question 7, soit

$$\widehat{Y}_{2012,1} = C_{2012,1} - C_{2012,0} = 9331.898 - 6184 = 3147.898$$

Bon, pour obtenir ce montant, ça ne devrait pas être pas trop compliqué, car on a $\widehat{\gamma}$ et $\widehat{\alpha}_{2012}$. Donc ici

$$\exp[8.77688 - 0.04716 + \widehat{\beta}_1] = 3147.898, \text{ i.e. } \widehat{\beta}_1 = \log(3147.898) - 8.77688 + 0.04716$$

soit

$$\widehat{\beta}_1 = -0.6752298$$

**Question 14.** Donnez la prédiction du montant de paiements qui sera effectué en 2015 pour les sinistres survenus en 2011. [*Le calcul sera explicité dans la feuille.*]

Pour les sinistres de 2011, payé en 2011, il faudrait $Y_{2011,0}$. Pour les payements faits en 2015, il faut aller chercher $Y_{2011,4}$, soit

$$\exp[\widehat{\gamma} + \widehat{\alpha}_{2011} + \widehat{\beta}_4] = \exp[8.77688 - 0.13771 - 3.83292] = 122.27$$

On notera que ce montant se retrouve aussi avec les calculs de la question 10,

$$Y_{2011,4} = C_{2011,4} - C_{2011,3} = C_{2011,2} \cdot \lambda_2 \cdot \lambda_3 \cdot [\lambda_4 - 1]$$

(je laisse vérifier ce peu calcul trivial) soit

$$Y_{2011,4} = 8524 \cdot \lambda_2 \cdot \lambda_3 \cdot [\lambda_4 - 1]$$

**Question 15.** En supposant la première année close, quel serait le montant de provision que l'on devrait constituer pour l'année de survenance 2010, à l'aide de cette régression ? [*Le calcul sera explicité dans la feuille.*]

Pour 2010, le montant de provision est donné particulier

$$Y_{2010,3} + Y_{2010,4} + Y_{2010,5} = 186.4313 + 122.4148 + 54.08676 = 362.93$$

On peut faire les calculs, et arriver au même montant que celui fait dans la question 11.

En cherchant à simplifier le modèle de manière à n'avoir que des variables significatives, on finit par considérer le modèle suivant

Régression 1b

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.74411    0.02632 332.270  < 2e-16 ***
i2010&11    -0.10544    0.04037  -2.612   0.0196 *
j1          -0.67226    0.04254 -15.802 9.29e-11 ***
j2          -2.03823    0.08161 -24.974 1.24e-13 ***
j3&4        -3.55337    0.14965 -23.745 2.59e-13 ***
j5          -4.61697    0.56126  -8.226 6.10e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for quasipoisson family taken to be 19.4877)
```

**Question 16.** Donnez la prédiction du montant de paiements qui sera effectué en 2015 pour les sinistres survenus en 2011. [*Le calcul sera explicité dans la feuille.*]

Pour les payements faits en 2015 pour les sinistres survenus en 2011, il faut aller chercher $Y_{2011,4}$ (comme dans la question 14). Ici, la prédiction sera

$$\exp[\widehat{\gamma} + \widehat{\alpha}_{2010\&11} + \widehat{\beta}_{3\&4}] = \exp[8.74411 - 0.10544 - 3.55337] = 161.6284.$$

**Question 17** [$\times 2$]**.** Calculez les montants de provisions induits par ce modèle [*Aucun calcul ne sera détaillé sur la feuille*]

Il faut calculer les $\widehat{Y}_{i,j}$ pour la partie inférieure du triangle. Mais les calculs sont ici assez simple car pour plusieurs endroits dans le triangle, les valeurs sont exactement les mêmes (comme les années de développement 3 et 4), ou avec juste deux années à considérer (2010&11 et les autres).

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 2007 | | | | | | |
| 2008 | | | | | | 62.0 |
| 2009 | | | | | 179.6 | 62.0 |
| 2010 | | | | 161.6 | 161.6 | 55.8 |
| 2011 | | | 735.4 | 161.6 | 161.6 | 55.8 |
| 2012 | | 3203.0 | 817.2 | 179.6 | 179.6 | 62.0 |

Si on somme, on obtient un montant total de provisions de l'ordre de

$$62.0 \cdot 3 + 55.8 \cdot 2 + 179.6 \cdot 3 + 161.6 \cdot 2 + 3203 + 817.2 + 735.4 = 6238.523$$

(que l'on comparera avec le montant obtenu avec la méthode Chain Ladder, 6058.89).

Dans la suite des questions 18-19, on va utiliser la sortie de régression suivante, toujours basée sur le triangle 1b :

Régression 2

```
Call:
lm(formula = log(Y) ~ i2008 + j, data = B)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.7531     0.1071  81.765  < 2e-16 ***
i2008        -0.2611     0.1354  -1.928 0.072953 .
j1           -0.6772     0.1553  -4.361 0.000559 ***
j2           -2.0118     0.1658 -12.131 3.73e-09 ***
j3&4         -3.5515     0.1584 -22.421 5.99e-13 ***
j5           -4.6259     0.2778 -16.653 4.40e-11 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.2563 on 15 degrees of freedom
```

**Question 18.** Quelle est le montant que l'on pense payer, avec ce modèle, en 2013 pour les sinistres survenus en 2012 ? [*Le calcul sera explicité dans la feuille.*]

Cette fois, on fait une régression lognormale. La prédiction pour $Y_{i,j}$ sera ici

$$\widehat{Y}_{i,j} = \exp\left[\widehat{\gamma} + \widehat{\alpha}_i + \widehat{\beta}_j + \frac{\sigma^2}{2}\right]$$

On nous demande ici de prédire $Y_{2012,1}$, soit

$$\exp\left[\widehat{\gamma} + \widehat{\beta}_1 + \frac{\sigma^2}{2}\right] = \exp\left[8.7531 - 0.6772 + \frac{0.2563^2}{2}\right] = \exp\left[8.108745\right] = 3323.405$$

(que l'on comparera avec le montant obtenu avec la méthode Chain Ladder, 3147.898).

**Question 19** [×2]**.** En supposant la première année close, quel serait le montant de provision que l'on devrait constituer à l'aide de cette régression, pour toutes les années ? [*Le calcul sera explicité dans la feuille.*]

Cette fois, il faut faire un peu de calcul, mais là encore, beaucoup seront identiques,

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 2007 | | | | | | |
| 2008 | | | | | | 49.3 |
| 2009 | | | | | 187.6 | 64.1 |
| 2010 | | | | 187.6 | 187.6 | 64.1 |
| 2011 | | | 874.9 | 187.6 | 187.6 | 64.1 |
| 2012 | | 3323.4 | 874.9 | 187.6 | 187.6 | 64.1 |

et pour avoir le montant total de provisions, on somme,

$$3323.4 + 874.9 \cdot 2 + 187.6 \cdot 7 + 64.1 \cdot 4 + 49.3 = 6692.311$$

(que l'on comparera avec le montant obtenu avec la méthode Chain Ladder, 6058.89).

## 2. Tarification

Trois exercices de tarifications sont proposés. Ils peuvent se traîter indépendamment. Il y a toutefois une logique à les faire dans l'ordre.

2.1. **Tarification, partie 1.** On suppose que pour un assuré, sa fréquence annuelle de sinistres suit une loi de Poisson de moyenne $\lambda$. Toutefois, on sait que la moitié des sinistres seulement dépasseront la franchise $d$. On sait aussi que si un sinistre dépasse $d$, il aura pour coût $Y$ de distribution conditionnelle

$$\mathbb{P}(Y > y|Y > d) = \left(\frac{d}{x}\right)^{\alpha}, \text{ pour } y \geq d, \text{ et sinon } 1.$$

**Question 20** [×2]**.** Quelle sera la prime pure d'un tel contrat ?

L'indemnité versée est ici

$$\sum_{i=1}^{N}(Y_i - d)_+ = \sum_{i=1}^{N}(Y_i - d)\mathbf{1}(Y_i > d) = \sum_{j=1}^{N_d}(Y_j - d)$$

où dans le terme de droite, la sommation se fait sur les sinistres qui ont coûté plus que $d$. On a alors deux solutions pour valoriser ce contrat (en croisant les doigts pour que ces deux méthodes coïncident).

**Méthode 1**. La prime pure sera

$$\mathbb{E}(N) \cdot \mathbb{E}[(Y_i - d)\mathbf{1}(Y_i > d)]$$

où $(Y - d)\mathbf{1}(Y > d)$ est une variable aléatoire qui vaut 0 si $Y \leq d$ et $Y - d$ si $Y > d$. Donc

$$\mathbb{E}[(Y - d)\mathbf{1}(Y > d)] = \mathbb{P}(Y > d) \cdot \mathbb{E}[(Y - d)|Y > d]$$

Or la loi en excès de $d$ est ici une loi de Pareto,

$$\mathbb{P}((Y - d) > x|Y > d) = \left(\frac{d}{d + x}\right)^{\alpha} \text{ pour } x > d,$$

de telle sorte que l'espérance s'écrit

$$\mathbb{E}[(Y - d)|Y > d] = \int_0^{\infty}\left(\frac{d}{x + d}\right)^{\alpha} dy = \left[d^{\alpha}\frac{(x + d)^{-\alpha+1}}{-\alpha + 1}\right]_0^{\infty}$$

soit

$$\mathbb{E}[(Y - d)|Y > d] = \frac{d}{\alpha - 1}$$

(si $\alpha > 1$, cf cours). Bref, ici la prime pure sera

$$\lambda \cdot \underbrace{\mathbb{P}(Y > d)}_{\frac{1}{2}} \cdot \frac{d}{\alpha - 1} = \frac{\lambda \cdot d}{2(\alpha - 1)}.$$

**Méthode 2**. La prime pure sera

$$\mathbb{E}(N_d) \cdot \mathbb{E}[(Y-d)|Y>d]$$

Comme rappelé au début du dernier cours, un propriété du Processus de Poisson est que $N_d$ suit une loi de Poisson, de paramètre $\lambda \cdot \mathbb{P}(Y>d)$, et donc la prime pure sera

$$\lambda \cdot \mathbb{P}(Y>d) \cdot \mathbb{E}[(Y-d)|Y>d]$$

On retrouve le calcul précédant.

**Question 21** [×2]**.** Afin d'attirer de nouveau client, on souhaite augmenter la franchise $d$ de manière à proposer un contrat deux fois moins cher. À quel niveau faudra-t-il placer la nouvelle franchise ?

Si on utilise une franchise $d$ plus élevée, disons $d'$, la formule évoquée auparavant sera encore valide, au sens où la prime pure (avec cette nouvelle franchise) sera

$$\lambda \cdot \mathbb{P}(Y>d') \cdot \mathbb{E}[(Y-d')|Y>d']$$

Reste à faire les calculs. En utilisant la formule des probabilités totales

$$\mathbb{P}(Y>d') = \mathbb{P}(Y>d'|Y>d) \cdot \mathbb{P}(Y>d) + \mathbb{P}(Y>d'|Y\leq d) \cdot \mathbb{P}(Y\leq d)$$

soit, puisque par hypothèse on souhaite monter la prime,

$$\mathbb{P}(Y>d') = \mathbb{P}(Y>d'|Y>d) \cdot \mathbb{P}(Y>d) + 0 = \left(\frac{d}{d'}\right)^{\alpha} \cdot \frac{1}{2}$$

De plus, compte tenu du calcul précédant,

$$\mathbb{P}[(Y-d')>x|Y>d'] = \mathbb{P}(Y>d'+x|Y>d) \cdot \mathbb{P}(Y>d'|Y>d) = \left(\frac{d}{d'+x}\right)^{\alpha} \cdot \left(\frac{d}{d'}\right)^{-\alpha}$$

en utilisant la formule de Bayes, soit

$$\mathbb{P}[(Y-d')>x|Y>d'] = \left(\frac{d'}{d'+x}\right)^{\alpha} \text{ pour } x>0,$$

On retrouve la loi de Pareto, de paramètre $\alpha$, mais avec comme second paramètre $d'$ et non plus $d$.

$$\mathbb{E}(Y-d'|Y>d') = \frac{d'}{\alpha-1}$$

On retrouve une formule évoquée en cours. Donc ici, la prime pure sera

$$\lambda \cdot \mathbb{P}(Y > d') \cdot \mathbb{E}[(Y - d')|Y > d'] = \lambda \cdot \left(\frac{d}{d'}\right)^{\alpha} \cdot \frac{1}{2} \cdot \frac{d'}{\alpha - 1}$$

Si on veut un contrat deux fois moins cher, il faudra question

$$\lambda \cdot \left(\frac{d}{d'}\right)^{\alpha} \cdot \frac{1}{2} \cdot \frac{d'}{\alpha - 1} = \frac{\lambda \cdot d}{2(\alpha - 1)} \cdot \frac{1}{2}$$

soit

$$\left(\frac{d}{d'}\right)^{\alpha} \cdot \frac{d'}{\alpha - 1} = \frac{d}{\alpha - 1} \cdot \frac{1}{2}$$

ou encore

$$\left(\frac{d}{d'}\right)^{\alpha - 1} = \frac{1}{2} \text{ i.e. } \frac{d'}{d} = 2^{\frac{1}{\alpha - 1}}$$

soit $d' = d \cdot 2^{\frac{1}{\alpha-1}}$ (avec $\alpha > 1$ car on souhaite calculer des primes pures).

2.2. **Tarification, partie 2.** Dans la section suivante, on cherche à construire un tarif en assurance automobile, partir de 2 variables explicative: l'âge d'un assuré et son sexe. On dispose d'un ensemble de données et plusieurs modèles sont ajustés. Tous les assurés sont restés un an dans la base. On suppose de plus quand les variables sexe et âge sont indépendantes. On suppose aussi disposer des statistiques suivantes

| | total | homme | femme |
|---|---|---|---|
| nombre moyen d'accident, par an | 0.0793 | 0.0838 | 0.07242 |
| coût moyenne d'un accident | 579.669 | 622.4477 | 504.2472 |
| moyenne du log(coût) d'un accident | 6.173865 | 6.249597 | 6.040344 |

Régression A

```
Call:
glm(formula = N ~ age + sexe, family = poisson)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.111191   0.127413 -16.570  < 2e-16 ***
age         -0.010135   0.002281  -4.442 8.89e-06 ***
sexeH       [redacted]  [redacted]  2.069   0.0385 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)
```

Régression B

```
Call:
glm(formula = N ~ cut(age, c(0, 30, 100)) + sexe, family = poisson,
    data = basen)

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)             -2.37019    0.11272 -21.027  < 2e-16 ***
age (30,100]            -0.28704    0.10988  -2.612  0.00899 **
sexeH                    0.14917    0.07391   2.018  0.04356 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)
```

Régression C

```
Call:
lm(formula = log(Y) ~ sexe, data = basec)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.04034    0.03601 167.734  < 2e-16 ***
sexeH        [redacted]   [redacted]    4.642 4.05e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6101 on 791 degrees of freedom
Multiple R-squared: 0.02651,        Adjusted R-squared: 0.02528
F-statistic: 21.54 on 1 and 791 DF,  p-value: 4.046e-06
```

Régression D

```
Call:
glm(formula = Y ~ sexe, family = Gamma(link = "log"), data = basec)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.22307    0.03791 164.146  < 2e-16 ***
sexeH        0.21059    0.04746   4.437 1.04e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for Gamma family taken to be 0.4125056)
```

On suppose que plusieurs compagnies sont en compétition sur le marché.

- compagnie 1 : pas de segmentation
- compagnie 2 : modèle A pour les nombres, C pour les coûts

- compagnie 3 : modèle B pour les nombres, C pour les coûts
- compagnie 4 : modèle A pour les nombres, D pour les coûts
- compagnie 5 : modèle B pour les nombres, D pour les coûts

**Question 22** [$\times$4]**.** Un jeune homme de 20 ans souhaite s'assurer. Quelle compagnie lui propose la prime pure la plus faible ? Quel est le montant proposé ? [*Les calculs seront détaillés dans la feuille.*]

Pour commencer, évaluons les valeurs manquantes dans les deux régressions. Pour la régression A, on *sait* (c'est dans le cours) que les prédictions par sous-groupes dans une régression de Poisson de Poisson coïncident avec les fréquences empiriques.

Prenons les personnes d'âge $x$. Le ratio de la fréquence pour les hommes et pour les femmes est ici

$$\frac{\exp[\widehat{\beta}_0 + \widehat{\beta}_1 \cdot x + \widehat{\beta}_2 \mathbf{1}(H)]}{\exp[\widehat{\beta}_0 + \widehat{\beta}_1 \cdot x + \widehat{\beta}_2 \mathbf{1}(F)]} = \exp[\widehat{\beta}_2]$$

où $\widehat{\beta}_2$ est précisément la valeur que l'on cherche, qui a été noircie. Si l'âge était parfaitement uniforme, et indépendant du sexe, ce ratio doit coïncider avec le ratio des fréquences, des hommes et des femmes, soit ici

$$\exp[\widehat{\beta}_2] \sim \frac{0.0838}{0.07242} = 1.157139$$

soit, en prenant le logarithme, $\widehat{\beta}_2 \sim \log(1.157139) = 0.1459$. Comme nous n'avons pas les âge, il était impossible de calculer exactement les prédictions. Mais un ordre de grandeur ici suffit.

Pour la régression C, c'est la même chose, car on considère ici une regression linéaire. C'est même encore plus simple car l'âge n'intervient plus La constante est donc le log-coût moyen pour les femmes ( effectivement, si on regarde dans la table, on retrouve une moyenne de 6.04034), et la somme des deux coefficients doit correspondre à la moyenne pour les hommes. Ici, on veut $\widehat{\beta}$ tel que

$$6.04034 + \widehat{\beta} = 6.249597$$

soit $\widehat{\beta} = 6.249597 - 6.04034 = 0.2093$.

Bon, cette fois, on a tout. Pour un homme de 20 ans, on a

- modèle A, fréquence prédite $\exp[-2.11191 - 0.010135 \cdot 20 + 0.1459] = 0.114325$,
- modèle B, fréquence prédite $\exp[-2.37019 + 0 + 0.14917] = 0.1084984$,
- modèle C, coût moyen prédit $\exp[6.04034 + 0.2093 + 0.6101^2/2] = 623.7508$,
- modèle D, coût moyen prédit $\exp[6.22307 + 0.21059] = 622.4479$,

Reste à combiner toutes ces valeurs pour avoir les montants de primes pures affichées par les différentes compagnies

- compagnie 1, $0.0793 \cdot 579.669 = 45.96775$,
- compagnie 2, $0.114325 \cdot 623.7508 = 71.310$
- compagnie 3, $0.1084984 \cdot 623.7508 = 67.67596$
- compagnie 4, $0.114325 \cdot 622.4479 = 71.16136$
- compagnie 5, $0.1084984 \cdot 622.4479 = 67.5346$

Il choisira donc la compagnie 1 (qui ne segmente pas). Les hommes et les jeunes était les plus risqués (signes des coefficients dans les régressions), on aurait pu avoir cette conclusion sans aucun calcul).

**Question 23** [$\times 2$]**.** Il trouve la prime trop chère, et demande à sa mère (qui a 50 ans) si elle peut assurer la voiture à son nom. Quelle compagnie proposera la prime pure la plus faible ? Quel est le montant proposé ? [*Les calculs seront détaillés dans la feuille.*]

Ce n'est pas très méchant comme question, il faut juste reprendre (rapidement) les derniers calculs Pour une femme de 50 ans, on a

- modèle A, fréquence prédite $\exp[-2.11191 - 0.010135 \cdot 50] = 0.07290048$,
- modèle B, fréquence prédite $\exp[-2.37019 - 0.28704 \cdot 1 + 0] = 0.07014225$,
- modèle C, coût moyen prédit $\exp[6.04034 + 0.6101^2/2] = 505.9567$,
- modèle D, coût moyen prédit $\exp[6.22307] = 504.2489$,

Reste à combiner toutes ces valeurs pour avoir les montants de primes pures affichées par les différentes compagnies

- compagnie 1, $0.0793 \cdot 579.669 = 45.96775$,
- compagnie 2, $0.07290048 \cdot 505.9567 = 36.88449$

- compagnie 3, $0.07014225 \cdot 505.9567 = 35.48894$
- compagnie 4, $0.07290048 \cdot 504.2489 = 36.75999$
- compagnie 5, $0.07014225 \cdot 504.2489 = 35.36915$

Sa mère choisira donc une compagnie qui propose de segmenter, la compagnie 5 étant ici la moins chére.

2.3. **Tarification, partie 3.**

Une autre compagnie arrive sur le marché, et propose d'utiliser une **loi exponentielle** pour les coûts de sinistres

Régression E

```
Call:
glm(formula = Y ~ sexe, family = Gamma(link = "log"), data = basec)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.22307    0.05903  105.42  < 2e-16 ***
sexeH        0.21059    0.07390    2.85  0.00437 **
---

(Dispersion parameter for Gamma family taken to be 1)
```

On a alors

- compagnie 6 : modèle A pour les nombres, E pour les coûts

**Question 24.** Un homme de 30 ans souhaite s'assurer. Quelle serait sa prime pure ? [*Les calculs seront détaillés dans la feuille.*]

La loi exponentielle est la loi Gamma, en forçant le paramètre de surdispersion à valoir 1. On a donc la même prédiction qu'avec la loi Gamma. Cette nouvelle compagnie propose le même prix que la compagnie 4. Bon, ici l'assuré a 30 ans, sa prime pure est alors

$$\exp[-2.11191 - 0.010135 \cdot 30 + 0.1459] \cdot \exp[6.22307 + 0.21059] = \exp[4.1636] = 64.3026.$$

**Question 25** [$\times 4$]**.** Il trouve la prime trop chère, mais cette fois, l'assureur lu propose un contrat avec une franchise de 100. En supposant les coûts et la fréquence indépendants, quelle sera la prime pure proposée ? [*Les calculs seront détaillés dans la feuille.*]

On a ici un contrat avec franchise. Si on pour les coûts une loi exponentielle de moyenne $\mu$ (ce qui est le cas ici)

$$\mathbb{P}(Y > y) = \exp[-y/\mu] \text{ si } y > 0.$$

Aussi (on retrouve une propriété centrale du cours de processus),

$$\mathbb{P}(Y - d > y | Y > d) = \frac{\mathbb{P}(Y - d > y)}{\mathbb{P}(Y > d)} = \frac{\exp[-(y + d)/\mu]}{\exp[-d/\mu]} = \exp[-y/\mu] \text{ si } y > 0.$$

Bref, on retrouve la même loi exponentielle: la loi exponentielle est sans mémoire (quand on l'utilise sur des durées).

En utilisant la question 20, on avait vu que la prime pure d'un contrat avec franchise serait

$$\lambda \cdot \mathbb{P}(Y > d) \cdot \mathbb{E}[(Y - d)|Y > d]$$

où $\lambda$ est la fréquence moyenne de sinistres. Donc ici, avec une loi de Poisson de paramètre $\lambda$ et une loi exponentielle de moyenne $\mu$, la prime pure d'un contrat avec franchise serait

$$\lambda \cdot \exp[-d/\mu] \cdot \mu,$$

soit $\exp[-d/\mu]$ fois la prime pure sans franchise, avec ici $\mu = \exp(6.22307 + 0.21059) = 622.4479$, soit $\exp[-100/622,4479] = 0.851585$. La nouvelle prime pure serait 15% moins chere que de celle proposé initialement, soit 54.75913.