

Équipe no XX

Nicholas Langevin
(111 184 631)

Alexandre Turcotte
(111 172 613)

Mathématiques actuarielles IARD 1
ACT-2005

Travail pratique 2

Travail présenté à
Andrew Luong

École d'actuariat
Université Laval
Automne 2018

Question 1

Méthode de vraisemblance avec données complètes

a) Estimation des paramètres

Le jeu de données utilisé afin de compléter ce numéro comportait les 10 données suivantes :

1500 6000 3500 3800 1800 5500 4800 4200 3900 3000

À partir de ces données et de SAS, il a été possible d'estimer les paramètres $\hat{\mu}$ et $\hat{\sigma}$ d'une loi log-normale à partir de la méthode du maximum de vraisemblance. D'ailleurs, le code SAS se trouve dans l'Annexe à la fin du rapport. Voici, les estimés des paramètres obtenus:

$\hat{\mu}$	$\hat{\sigma}$
8.1618	0.4278

Table 1: Paramètres estimés d'une loi log-normale à l'aide du maximum de vraisemblance

Il est également possible de déterminer les estimateurs de façon théorique. D'abord, il faut déterminer la fonction $g(x_i; \mu, \sigma)$ pour les dix scénarios, et ce, en fonction de la densité de X.

Alors comme $X \sim LN(\mu, \sigma)$, la densité de f(x) est la suivante :

$$f_x(x) = \frac{1}{\sqrt{2\pi x \sigma}} e^{-\frac{1}{2} \left(\frac{\ln(x) - \mu}{\sigma} \right)^2}$$

Ensuite, il faut appliquer la vraisemblance en tenant compte des 10 scénarios et de la densité de X.

$$\begin{aligned} \mathbb{L}(\mu, \sigma) &= \prod_{i=1}^{10} g(x_i; \mu, \sigma) \\ &= f_x(1500) \cdot f_x(6000) \cdot \dots \cdot f_x(3000) \\ &= \left[\frac{1}{\sqrt{2\pi 1500 \sigma}} e^{-\frac{1}{2} \left(\frac{\ln(1500) - \mu}{\sigma} \right)^2} \right] \cdot \left[\frac{1}{\sqrt{2\pi 6000 \sigma}} e^{-\frac{1}{2} \left(\frac{\ln(6000) - \mu}{\sigma} \right)^2} \right] \cdot \dots \cdot \left[\frac{1}{\sqrt{2\pi 3000 \sigma}} e^{-\frac{1}{2} \left(\frac{\ln(3000) - \mu}{\sigma} \right)^2} \right] \\ &= \left(\frac{1}{2\pi} \right)^5 \left(\frac{1}{\sigma} \right)^{10} \left(\frac{1}{2.795153962 \times 10^{35}} \right) e^{-\frac{1}{2\sigma^2} [(\ln(1500) - \mu)^2 + (\ln(6000) - \mu)^2 + \dots + (\ln(3000) - \mu)^2]} \end{aligned}$$

Par la suite, il faut appliquer la log vraisemblance.

$$\begin{aligned}
l(\mu, \sigma) &= \ln(L(\mu, \sigma)) = \ln \prod_{i=1}^{10} g(x_i; \mu, \sigma) \\
&= -5\ln(2\pi) - 10\ln(\sigma) - \ln(2.795153962 \times 10^{35}) \\
&\quad - \frac{1}{2\sigma^2} [(\ln(1500) - \mu)^2 + (\ln(6000) - \mu)^2 + \dots + (\ln(3000) - \mu)^2]
\end{aligned}$$

Finalement, il faut dériver la log vraisemblance selon les deux paramètres et égaliser les deux équations à 0 pour pouvoir estimer les paramètres en les isolant.

Dériver selon μ

$$\begin{aligned}
\frac{d}{d\mu} l(\mu, \sigma) &= \frac{-1}{2\sigma^2} [-2(\ln(1500) - \mu) - 2(\ln(6000) - \mu) - \dots - 2(\ln(3000) - \mu)] \\
0 &= \frac{1}{\hat{\sigma}^2} [\ln(1500) + \ln(6000) + \dots + \ln(3000) - 10\hat{\mu}] \\
\hat{\mu} &= \frac{1}{10} \sum_{i=1}^{10} \ln(x_i) = 8.161836544
\end{aligned}$$

Dériver selon σ

$$\begin{aligned}
\frac{d}{d\sigma} l(\mu, \sigma) &= \frac{-10}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{10} (\ln(x_i) - \mu)^2 \\
0 &= \frac{-10}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} \sum_{i=1}^{10} (\ln(x_i) - \hat{\mu})^2 \\
\hat{\sigma}^2 &= \frac{1}{10} \sum_{i=1}^{10} (\ln(x_i) - 8.161836544)^2 = 0.1830137502 \\
\hat{\sigma} &= \sqrt{\hat{\sigma}^2} = 0.427801064
\end{aligned}$$

Donc, les estimateurs de la loi log-normale, selon la méthode du maximum de vraisemblance, sont $\hat{\mu} = 8.1618$ et $\hat{\sigma} = 0.4278$.

b) Détermination des primes stop-loss

En ayant obtenu les estimés des paramètres à la question a), il est maintenant possible d'effectuer divers calculs puisque la distribution du modèle est maintenant connue. Par conséquent, il est possible de calculer la prime stop-loss $E[(X - d)_+]$, et ce, pour différentes valeurs de déductible, soit $d = 2000, 2100, \dots, 3000$. Pour y arriver, il faut utiliser la formule suivante:

$$E[(X - d)_+] = e^{\hat{\mu} + \frac{\hat{\sigma}^2}{2}} \left(1 - \Phi \left(\frac{Ind - \hat{\mu} - \hat{\sigma}^2}{\hat{\sigma}} \right) \right) - d \left(1 - \Phi \left(\frac{Ind - \hat{\mu}}{\hat{\sigma}} \right) \right)$$

où $\hat{\mu} = 8.1618$, $\hat{\sigma} = 0.4278$ et $d =$ déductible. Par conséquent, les valeurs obtenues pour la prime stop-loss sont les suivantes:

d	Prime stop-loss
2000	1872.5736
2100	1783.0847
2200	1695.7633
2300	1610.7845
2400	1528.3012
2500	1448.4425
2600	1371.3131
2700	1296.9934
2800	1225.5400
2900	1156.9871
3000	1091.3484

Table 2: Valeurs de la prime stop-loss pour chacun des déductibles d donnés

À partir des valeurs de la *Table 2*, il est possible d'observer une baisse du niveau de la prime lorsque le déductible augmente. De plus, en regardant le graphique (A) de la *Figure 1*, il est possible d'observer cette décroissance qui semble à première vue assez linéaire. Toutefois, en regardant le graphique (B) de la *Figure 1*, ayant un plus grand nombre de valeurs de d , on remarque la convexité de la courbe, ce qui signifie que plus d augmente, moins la diminution est significative. Par exemple, l'écart entre 2000 et 2100 est plus important que l'écart entre 2900 et 3000. Bref, cela fait du sens puisqu'en ajoutant un déductible, la compagnie vient diminuer le risque ainsi que le montant de prestation qu'elle devra verser. Par conséquent, le niveau de la prime devrait diminuer, et ce, avec un impact de moins en moins significatif, car la valeur ajoutée d'augmenter le déductible pour la compagnie est de moins en moins importante.

c) Détermination de l'espérance limitée

Maintenant, toujours en utilisant les paramètres estimés en a), il faut désormais calculer l'espérance limitée, sans déductible, et ce, pour différentes limites supérieures, soit $u = 3000, 3100, \dots, 4000$. Pour y arriver, il faut utiliser la formule suivante:

$$E[X \wedge u] = e^{\hat{\mu} + \frac{\hat{\sigma}^2}{2}} \Phi \left(\frac{In(u) - \hat{\mu} - \hat{\sigma}^2}{\hat{\sigma}} \right) + u \left(1 - \Phi \left(\frac{In(u) - \hat{\mu}}{\hat{\sigma}} \right) \right)$$

où $\hat{\mu} = 8.1618$, $\hat{\sigma} = 0.4278$ et $u =$ limite. Par conséquent, les valeurs obtenues pour l'espérance limitée sont les suivantes:

u	Espérance limitée
3000	2748.9541
3100	2811.6843
3200	2871.5287
3300	2928.5251
3400	2982.7242
3500	3034.1882
3600	3082.9884
3700	3129.2039
3800	3172.9195
3900	3214.2249
4000	3253.2129

Table 3: Valeurs de l'espérance limitée pour chacune des limites u données

À partir des valeurs de la *Table 3*, il est possible d'observer une hausse du niveau de la prime lorsque la limite augmente. De plus, en regardant le graphique (C) de la *Figure 1*, il est possible d'observer cette croissance. D'ailleurs, en regardant le graphique (D) de la *Figure 1*, ayant un plus grand nombre de valeurs u , on remarque la concavité de la courbe, ce qui signifie que plus u augmente, moins l'augmentation est significative. Bref, cela fait du sens puisqu'en ayant une limite plus faible, la compagnie vient diminuer le risque ainsi que le montant de prestation qu'elle devra verser puisqu'elle ne payera pas ce qui a en excédent. Par conséquent, le niveau de la prime va augmenter plus la limite est grande, car la compagnie devra déboursier de plus grandes sommes, et ce, avec un impact de moins en moins significatif, car, à un certain point, le fait d'augmenter la limite va affecter peu de gens, donc la hausse des sommes à déboursier sera moins importante.

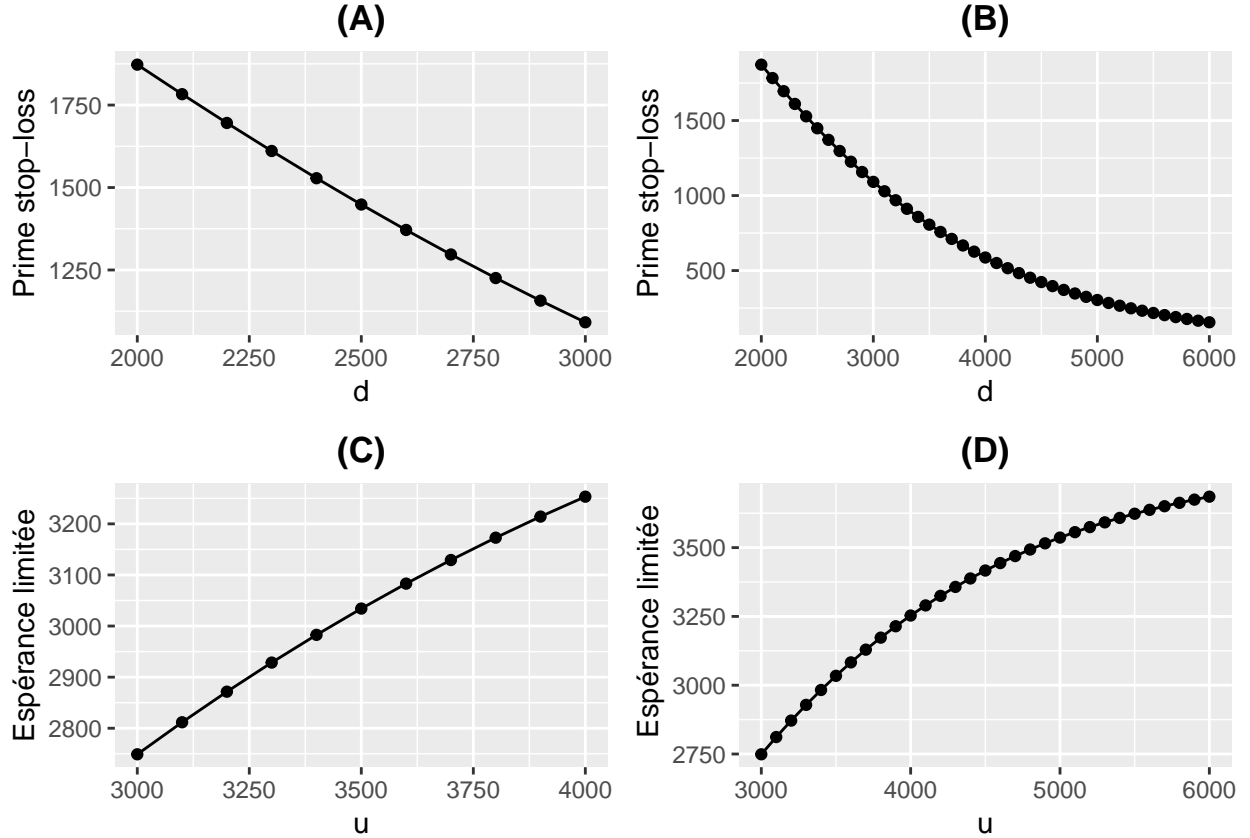


Figure 1 : (A) et (B) Représentation graphique de la prime stop-loss en fonction de d . (C) et (D) Représentation graphique de l'espérance limitée en fonction de u . Toutefois, l'étendu des valeurs sur l'axe des

abscisses est plus importante pour les graphiques (B) et (D) afin de mieux discerner la tendance.

d) Détermination de la prime stop-loss par simulation avec un déductible de 2000

Il est également possible d'approximer la prime stop-loss par simulation. Alors, il est possible de tirer un échantillon de la loi log-normale suffisamment grand pour estimer la prime stop-loss. Cela a été fait avec un $m = 100\,000$, $500\,000$ et $1\,000\,000$. Pour y arriver, il faut utiliser la formule suivante:

$$E[(X - 2000)_+] = \frac{1}{m} \sum_{i=1}^m \max(0, X - 2000)$$

Par conséquent, les valeurs obtenues pour la prime stop-loss sont les suivantes:

m	approximation
100000	1870.0455
500000	1872.1354
1000000	1872.3536

Table 4: Valeurs approximées par simulations de la prime stop-loss ayant un déductible de 2000

La valeur de la prime stop-loss calculée en b) pour un déductible de 2000 est de 1872.5736. Puis, en regardant les résultats obtenus à l'aide de la simulation, on se rend compte que ceux-ci sont très près de la valeur théorique, soit 1870.0455, 1872.1354 et 1872.3536 pour 100 000, 500 000 et 1 000 000 simulations respectivement. En effet, les écarts relatifs sont de 0.135%, 0.0234% et 0.0117%. Par conséquent, l'approximation est très précise et donne sensiblement la même valeur que la valeur théorique lorsqu'un grand nombre de simulations est effectué.

Question 2

Méthode de vraisemblance avec données groupées

a) Estimation des paramètres

Le jeu de données utilisé afin de compléter ce numéro comportait les données suivantes :

Intervalles	nombres
0-250	3
250-500	4
500-750	5
750-1000	5
1000-1250	4
1250-1500	3
1500-2000	3
2000-2500	3
2500-3000	2
3000-5000	2

Table 5: Jeux de données de la question 2

À partir de ces données et de SAS, il a été possible d'estimer les paramètres $\hat{\mu}$ et $\hat{\sigma}$ d'une loi log-normale à partir de la méthode du maximum de vraisemblance. D'ailleurs, le code SAS se trouve dans l'Annexe à la fin du rapport. Voici, les estimés des paramètres obtenus:

$\hat{\mu}$	$\hat{\sigma}$
6.8418	0.8278

Table 6: Paramètres estimés d'une loi log-normale à l'aide du maximum de vraisemblance

Afin de déterminer les estimateurs de la loi log-normale, il faut d'abord trouver la fonction $g(x_i; \mu, \sigma)$ pour l'ensemble des scénarios, et ce, en fonction de la répartition de X puisque les valeurs correspondent à un interval et non à une valeur ponctuelle.

Alors comme $X \sim LN(\mu, \sigma)$, la fonction de répartition $F(x)$ est la suivante :

$$F_x(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right)$$

Ensuite, il faut appliquer la vraisemblance en tenant compte de tous les scénarios possibles et de la répartition de X.

$$\begin{aligned} \mathbb{L}(\mu, \sigma) &= \prod_{i=1}^{10} g(x_i; \mu, \sigma) \\ &= [F_x(250) - F_x(0)]^3 \cdot [F_x(500) - F_x(250)]^4 \cdot \dots \cdot [F_x(5000) - F_x(3000)]^2 \end{aligned}$$

$$= \left[\Phi \left(\frac{In(250) - \mu}{\sigma} \right) - \Phi \left(\frac{In(0) - \mu}{\sigma} \right) \right]^3 \cdot \left[\Phi \left(\frac{In(500) - \mu}{\sigma} \right) - \Phi \left(\frac{In(250) - \mu}{\sigma} \right) \right]^4 \\ \cdot \dots \cdot \left[\Phi \left(\frac{In(5000) - \mu}{\sigma} \right) - \Phi \left(\frac{In(3000) - \mu}{\sigma} \right) \right]^2$$

Par la suite, il faut appliquer la log vraisemblance.

$$\begin{aligned} \ln(\mu, \sigma) &= \ln(L(\mu, \sigma)) = \ln \prod_{i=1}^{10} g(x_i; \mu, \sigma) \\ &= 3 \ln \left[\Phi \left(\frac{In(250) - \mu}{\sigma} \right) - 0 \right] + 4 \ln \left[\Phi \left(\frac{In(500) - \mu}{\sigma} \right) - \Phi \left(\frac{In(250) - \mu}{\sigma} \right) \right] \\ &\quad + \dots + 2 \ln \left[\Phi \left(\frac{In(5000) - \mu}{\sigma} \right) - \Phi \left(\frac{In(3000) - \mu}{\sigma} \right) \right] \end{aligned}$$

Finalement, il faut chercher à maximiser la fonction log vraisemblance afin d'obtenir les estimateurs $\hat{\mu}$ et $\hat{\sigma}$. Cela a été fait à l'aide du logiciel SAS.

Donc, les estimateurs de la loi log-normale, selon la méthode du maximum de vraisemblance, sont $\hat{\mu} = 6.8418$ et $\hat{\sigma} = 0.8278$.

b) Détermination des primes stop-loss

En ayant obtenu les estimés des paramètres à la question a), il est maintenant possible d'effectuer divers calculs puisque la distribution du modèle est maintenant connue. Par conséquent, il est possible de calculer la prime stop-loss $E[(X - d)_+]$, et ce, pour différentes valeurs de déductible, soit $d = 2000, 2100, \dots, 3000$. Pour y arriver, il faut utiliser la formule suivante:

$$E[(X - d)_+] = e^{\hat{\mu} + \frac{\hat{\sigma}^2}{2}} \left(1 - \Phi \left(\frac{Ind - \hat{\mu} - \hat{\sigma}^2}{\hat{\sigma}} \right) \right) - d \left(1 - \Phi \left(\frac{Ind - \hat{\mu}}{\hat{\sigma}} \right) \right)$$

où $\hat{\mu} = 6.8418$, $\hat{\sigma} = 0.8278$ et $d =$ déductible. Par conséquent, les valeurs obtenues pour la prime stop-loss sont les suivantes:

À partir des valeurs de la *Table 7*, il est possible d'observer une baisse du niveau de la prime lorsque le déductible augmente. De plus, en regardant le graphique (A) de la *Figure 2*, il est possible d'observer cette décroissance qui semble à première vue assez linéaire. Toutefois, en regardant le graphique (B) de la *Figure 2*, ayant un plus grand nombre de valeurs de d , on remarque la convexité de la courbe, ce qui signifie que plus d augmente, moins la diminution est significative. Bref, cela fait du sens puisqu'en ajoutant un déductible, la compagnie vient diminuer le montant qu'elle devra verser en prestations. Par conséquent, le niveau de la prime devrait diminuer, et ce, avec un impact de moins en moins significatif, car la valeur ajoutée d'augmenter le déductible pour la compagnie est de moins en moins importante.

c) Détermination de l'espérance limitée

Maintenant, toujours en utilisant les paramètres estimés en a), il faut désormais calculer l'espérance limitée, sans déductible, et ce, pour différentes limites supérieure, soit $u = 3000, 3100, \dots, 4000$. Pour y arriver, il faut utiliser la formule suivante:

d	Prime stop-loss
2000	253.3579
2100	236.1653
2200	220.3997
2300	205.9213
2400	192.6061
2500	180.3436
2600	169.0355
2700	158.5939
2800	148.9402
2900	140.0041
3000	131.7223

Table 7: Valeurs de la prime stop-loss pour chacun des déductibles d donnés

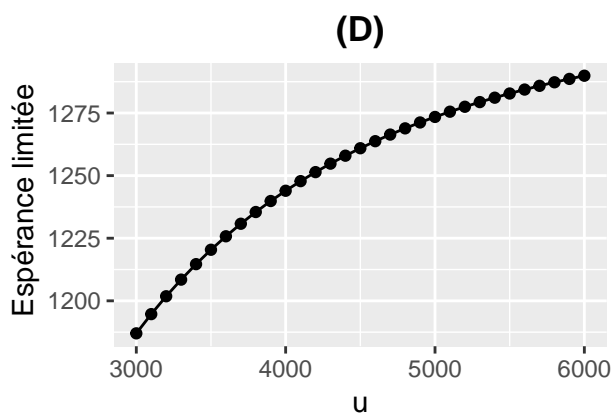
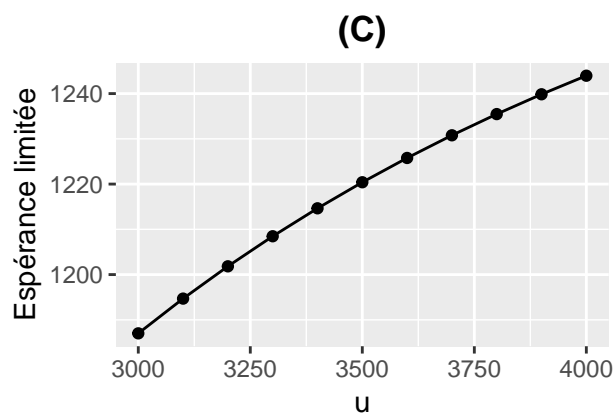
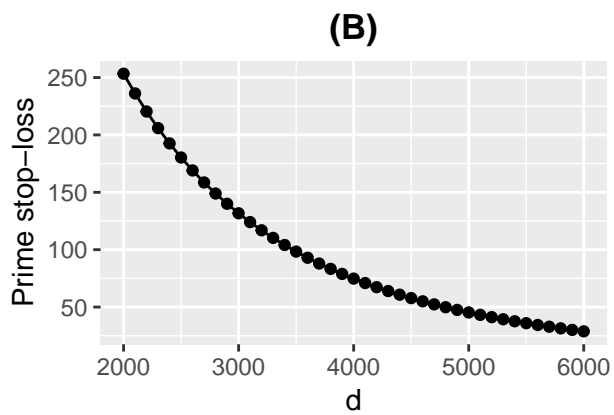
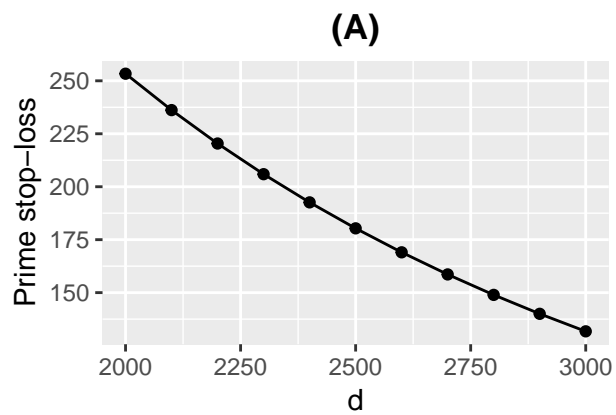
$$E[X \wedge u] = e^{\hat{\mu} + \frac{\hat{\sigma}^2}{2}} \Phi\left(\frac{Inu - \hat{\mu} - \hat{\sigma}^2}{\hat{\sigma}}\right) + u \left(1 - \Phi\left(\frac{Inu - \hat{\mu}}{\hat{\sigma}}\right)\right)$$

où $\hat{\mu} = 6.8418$, $\hat{\sigma} = 0.8278$ et $u = \text{limite}$. Par conséquent, les valeurs obtenues pour l'espérance limitée sont les suivantes:

u	Espérance limitée
3000	1187.0103
3100	1194.6945
3200	1201.8320
3300	1208.4691
3400	1214.6471
3500	1220.4038
3600	1225.7731
3700	1230.7858
3800	1235.4700
3900	1239.8511
4000	1243.9524

Table 8: Valeurs de l'espérance limitée pour chacune des limites u données

À partir des valeurs de la *Table 8*, il est possible d'observer une hausse du niveau de la prime lorsque la limite augmente. De plus, en regardant le graphique (C) de la *Figure 2*, il est possible d'observer cette croissance. D'ailleurs, en regardant le graphique (D) de la *Figure 2*, ayant un plus grand nombre de valeurs u , on remarque la concavité de la courbe, ce qui signifie que plus u augmente, moins l'augmentation est significative. Bref, cela fait du sens puisqu'en ayant une limite plus faible, la compagnie vient diminuer le risque ainsi que le montant de prestation qu'elle devra verser puisqu'elle ne payera pas ce qui a en excédent. Par conséquent, le niveau de la prime va augmenter plus la limite est grande, car la compagnie devra déboursier de plus grandes sommes, et ce, avec un impact de moins en moins significatif, car, à un certain point, le fait d'augmenter la limite va affecter peu de gens, donc la hausse des sommes à déboursier sera moins importante.



Question 3

Estimateur de Bayes

a) Estimation du paramètre θ

On trouve le maximum de vraisemblance pour la loi binomiale négative avec $r = 2$ et le paramètre θ , qui représente la probabilité de succès. On suppose de prime à bord qu'il n'y a pas de loi à priori. Le maximum de vraisemblance est alors défini par:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n \binom{r+x_i-1}{x_i} \theta^r (1-\theta)^{x_i} \\ &= \theta^{nr} (1-\theta)^{\sum_{i=1}^n x_i} \prod_{i=1}^n \binom{r+x_i-1}{x_i} \end{aligned}$$

On trouve alors la log-vraisemblance

$$\begin{aligned} l(\theta) &= \ln(\theta^{nr}) + \ln(1-\theta)^{\sum_{i=1}^n x_i} - \sum_{i=1}^n \ln \binom{r+x_i-1}{x_i} \\ &= nr \ln \theta + \sum_{i=1}^n x_i \ln(1-\theta) - \sum_{i=1}^n \ln \binom{r+x_i-1}{x_i} \end{aligned}$$

On dérive ensuite par rapport à θ pour trouver l'estimateur

$$\begin{aligned} \frac{dl(\theta)}{d\theta} &= \frac{nr}{\theta} - \frac{\sum_{i=1}^n x_i}{(1-\theta)} = 0 \\ \Rightarrow \frac{nr}{\hat{\theta}} &= \frac{\sum_{i=1}^n x_i}{(1-\hat{\theta})} \\ \Rightarrow \frac{nr(1-\hat{\theta})}{\hat{\theta}} &= \sum_{i=1}^n x_i \\ \Rightarrow \frac{nr}{\hat{\theta}} &= \sum_{i=1}^n x_i + nr \\ \Rightarrow \hat{\theta}_{MLE} &= \frac{nr}{\sum_{i=1}^n x_i + nr} \\ \Rightarrow \hat{\theta}_{MLE} &= \frac{r}{\bar{x} + r} \end{aligned}$$

On calcule ensuite le maximum de vraisemblance avec notre unique donnée $x = 5$

$$\hat{\theta}_{MLE} = \frac{r}{\bar{x} + r} = \frac{2}{5 + 2} = \frac{2}{7}$$

b) Présentation graphique de la loi beta a priori

On choisit comme loi à priori la loi beta avec $\alpha = 2$ et $\beta = 2$. On prend la loi beta puisque son domaine respecte le domaine de notre estimateur, soit $\theta \in [0, 1]$. La figure 1 permet d'observer la moyenne de la distribution (point rouge) qui est égal à 0.5. L'équation 3.1 permet de confirmer cette valeur.

$$\begin{aligned}\theta &\sim \text{Beta}(\alpha = 2, \beta = 2) \\ E[\theta] &= \frac{\alpha}{\alpha + \beta} = 0.5\end{aligned}\tag{3.1}$$

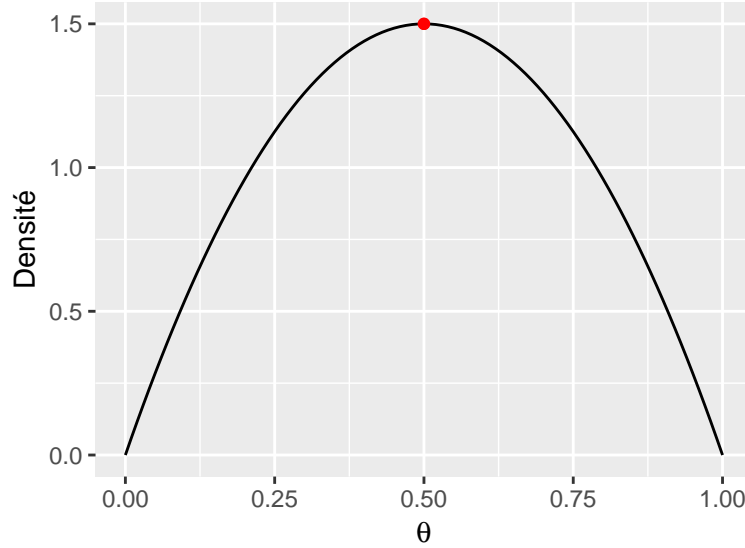


Figure 1: Densité d'une loi beta($\alpha = 2, \beta = 2$)

c) Détermination de la loi à posteriori ainsi que la valeur exacte de l'estimateur de Bayes

On détermine la loi à posteriori, pour ensuite trouver l'estimateur de Bayes pour notre θ . Puisque notre loi a priori est une Beta, la loi à posteriori devrait elle aussi suivre une Beta avec α' et β' .

$$\begin{aligned}f(\theta|x_1, \dots, x_n) &= \frac{f(\theta) \cdot f(\theta|x_1, \dots, x_n)}{\int f(\theta) \cdot f(\theta|x_1, \dots, x_n) d\theta} \\ &= c(x) \cdot \theta^{\alpha-1} (1-\theta)^{\beta-1} \theta^r (1-\theta)^{x_1} \dots \theta^r (1-\theta)^{x_n} \\ &= c(x) \cdot \theta^{\alpha+nr-1} (1-\theta)^{\beta+\left(\sum_{i=1}^n x_i\right)-1} \\ &\sim \text{Beta}(\alpha' = \alpha + nr, \beta' = \beta + \sum_{i=1}^n x_i)\end{aligned}\tag{3.2}$$

où $c(x)$ est une constante faisant en sorte que la densité intégrée sur tout son domaine égale à 1. Ensuite, on peut trouver l'estimateur de Bayes pour notre θ , et ce, en utilisant notre unique donnée $x = 5$.

$$\begin{aligned}
\hat{\theta}_B &= \int_0^1 \theta f(\theta|x) d\theta \\
&= \frac{\alpha'}{\alpha' + \beta'} \\
&= \frac{\alpha + nr}{\alpha + nr + \beta + \sum_{i=1}^n x_i} \\
&= \frac{\alpha}{\alpha + nr + \beta + \sum_{i=1}^n x_i} + \frac{rn}{\alpha + nr + \beta + \sum_{i=1}^n x_i} \\
&= \frac{2 + (1)2}{2 + (1)2 + 2 + 5} \\
&= \frac{4}{11}
\end{aligned} \tag{3.3}$$

L'estimateur de Bayes estime une probabilité plus grande pour notre θ que l'estimateur de vraisemblance. Puisque nous avons une seule donnée, l'estimateur du maximum de vraisemblance possède une plus grande variance que l'estimateur de Bayes. Par contre, Bayes détermine un estimateur biaisé. En utilisant la représentation en 3.3 et en prenant $n \rightarrow \infty$, il est possible de montrer que l'estimateur de Bayes tend vers l'estimateur du maximum de vraisemblance. On en déduit donc que l'estimateur de Bayes est préférable pour un petit échantillon de données, alors que celui du maximum de vraisemblance sera plus précis pour un grand échantillon.

$$\begin{aligned}
\lim_{n \rightarrow \infty} \hat{\theta}_B &= \frac{\alpha}{\alpha + nr + \beta + \sum_{i=1}^n x_i} + \frac{rn}{\alpha + nr + \beta + \sum_{i=1}^n x_i} \\
&= \lim_{n \rightarrow \infty} \frac{\alpha}{\alpha + nr + \beta + \sum_{i=1}^n x_i} + \lim_{n \rightarrow \infty} \frac{rn}{n(\frac{\alpha}{n} + r + \frac{\beta}{n} + \bar{x})} \\
&= 0 + \lim_{n \rightarrow \infty} \frac{r}{\frac{\alpha}{n} + r + \frac{\beta}{n} + \bar{x}} \\
&= \frac{r}{r + \bar{x}} = \hat{\theta}_{MLE}
\end{aligned}$$

d) Approximation de l'estimateur de Bayes par simulation

On simule notre estimateur de Bayes pour le comparer avec la valeur théorique calculée en c). Pour ce faire on effectue $m = 500000$ simulations de la loi à posteriori. Comme il a été démontré à l'équation 3.2, cette loi suit aussi une beta avec $\alpha = 4$ et $\beta = 7$. L'estimateur de Bayes simulé consiste alors à la moyenne de ces simulations. Avec cette méthode et ce nombre important de simulations, nous obtenons $\hat{\theta}_B = 0.3636007$. Le résultat obtenu est différent seulement à partir de la 5^e décimal, ce qui est très précis. Pour plus de précision, il aurait été possible d'augmenter le nombre de simulations, mais cela aurait requis plus de temps et potentiellement plus de puissance machine pour une énorme quantité de simulations.

Annexe

Question 1

a) Estimation des paramètres

```
# Données utilisées afin de compléter le numéro 1

data <- c(1500, 6000, 3500, 3800, 1800, 5500, 4800, 4200, 3900, 3000)

## [1] 1500 6000 3500 3800 1800 5500 4800 4200 3900 3000

# Code SAS permettant d'estimer les paramètres de la loi log-normale
# par la méthode du maximum de vraisemblance

data MLE;
input t cens;
datalines;
1500 1
6000 1
3500 1
3800 1
1800 1
5500 1
4800 1
4200 1
3900 1
3000 1
;
run;

proc lifereg data=MLE;
  Model t*cens(0)= /covb dist=lnormal;
run;
```

b) Détermination des primes stop-loss

```
# Les estimateurs ont été calculés à l'aide de SAS en a)
mu <- 8.1618
sigma <- 0.4278

# Valeurs des déductibles d à utiliser
d <- 100 * seq(20,30)

# Calculer les primes STOP-LOSS pour les différents déductibles d
STOP_LOSS <- function(x) {
  exp(mu + sigma^2/2) * (1 - pnorm((log(x) - mu - sigma^2)/sigma)) -
    x * (1 - pnorm((log(x) - mu)/sigma))
}
```

```

Primes_STOP_LOSS <- data.frame(d, "Prime" = sapply(d, function(x) STOP_LOSS(x)))

# Présenter les valeurs des primes STOP-LOSS en fonction de d
library(xtable)
options(xtable.comment = FALSE)
xtable(Primes_STOP_LOSS, caption = "Valeurs de la prime stop-loss pour
      chacun des déductibles d donnés",
      align = c("c", "c", "c"),
      digits = c(0, 0, 4))

# Représenter graphiquement l'évolution de la prime stop-loss en fonction de d
library(ggplot2)
plot1 <- ggplot(data = Primes_STOP_LOSS, aes(x= Primes_STOP_LOSS$d,
      y = Primes_STOP_LOSS$Prime))
  + geom_point() + geom_line() + xlab("d") + ylab("Prime stop-loss")
  + ggtitle("(A)") + theme(plot.title = element_text(face="bold", hjust = 0.5))

value_test <- data.frame(d=(100 * seq(20,60)), "Prime" = sapply((100 * seq(20,60)),
      function(x) STOP_LOSS(x)))

plot2 <- ggplot(data=value_test, aes(x=value_test$d, y=value_test$Prime))
  + geom_point() + geom_line() + xlab("d") + ylab("Prime stop-loss")
  + ggtitle("(B)") + theme(plot.title = element_text(face="bold", hjust = 0.5))

```

c) Détermination des primes limited loss

```

# Valeurs des limites u à utiliser
u <- 100 * seq(30,40)

# Calculer l'espérance limitée pour les différentes limites u
Esp_limite <- function(x) {
  exp(mu + sigma^2/2) * pnorm((log(x) - mu - sigma^2)/sigma) +
    x * (1 - pnorm((log(x) - mu)/sigma))
}

Primes_Esp_limite <- data.frame(u, "Prime" = sapply(u, function(x) Esp_limite(x)))

# Présenter les valeurs des espérances limitées en fonction de u
library(xtable)
options(xtable.comment = FALSE)
xtable(Primes_Esp_limite, caption = "Valeurs de l'espérance limitée pour
      chacune des limites u données",
      align = c("c", "c", "c"),
      digits = c(0, 0, 4))

# Représenter graphiquement l'évolution de l'espérance limitée en fonction de u
library(ggplot2)
plot3 <- ggplot(data = Primes_Esp_limite, aes(x= Primes_Esp_limite$u,
      y = Primes_Esp_limite$Prime))
  + geom_point() + geom_line() + xlab("u") + ylab("Espérance limitée") + ggtitle("(C)")
  + theme(plot.title = element_text(face="bold", hjust = 0.5))

```

```

value_test_2 <- data.frame(u=(100 * seq(30,60)), "Prime" = sapply((100 * seq(30,60)),
  function(x) Esp_limite(x)))

plot4 <- ggplot(data=value_test_2, aes(x=value_test_2$u, y=value_test_2$Prime))
  + geom_point() + geom_line() + xlab("u") + ylab(" Espérance limitée")
  + ggtitle("(D)") + theme(plot.title = element_text(face="bold", hjust = 0.5))

# Mettre les 4 graphiques ensembles
library(gridExtra)
grid.arrange(plot1, plot2, plot3, plot4, ncol=2, nrow=2)

```

d) Détermination de la prime stop-loss par simulation avec un déductible $d = 2000$

```

# Valeurs du déductible d à utiliser
d <- 2000

# Nombre de simulations
m <- c(1e5, 5e5, 1e6)

# Simuler un échantillon de m simulations de la log log-normale
# avec les paramètres estimés en a)
simul <- function(x) rlnorm(x, meanlog = mu, sdlog = sigma)
simulation <- sapply(1:3, function(i) simul(m[i]))

# Approximer la prime STOP-LOSS ayant un déductible de 2000 par simulation
approximation <- sapply(1:3, function(x)
  mean(sapply(1:m[x], function(i)
    max(0, simulation[[x]][i] - 2000))))

approx <- data.frame(m, approximation)

# Trouver les écarts relatifs (par rapport à la valeur théorique)
Prime_2000 <- Primes_STOP_LOSS$Prime[1]
Approx <- approx$approximation

Ecart <- abs((Approx-Prime_2000)/Prime_2000) * 100

# Présenter les valeurs approximées de la prime STOP-LOSS ayant un déductible de 2000
library(xtable)
options(xtable.comment = FALSE)
xtable(approx, caption = "Valeurs approximées par simulations de la prime stop-loss
  ayant un déductible de 2000",
  align = c("c", "c", "c"),
  digits = c(0, 0, 4))

```


Question 2

a) Estimation des paramètres

```
# Code SAS permettant d'estimer les paramètres de la loi log-normale  
# par la méthode du maximum de vraisemblance  
data donnees;  
input lb ub;  
datalines;  
0.00001 250  
0.00001 250  
0.00001 250  
250 500  
250 500  
250 500  
250 500  
500 750  
500 750  
500 750  
500 750  
500 750  
750 1000  
750 1000  
750 1000  
750 1000  
750 1000  
1000 1250  
1000 1250  
1000 1250  
1000 1250  
1250 1500  
1250 1500  
1250 1500  
1500 2000  
1500 2000  
1500 2000  
2000 2500  
2000 2500  
2000 2500  
2500 3000  
2500 3000  
3000 5000  
3000 5000  
;  
run;  
  
proc lifereg data=donnees;  
    Model (lb,ub)= /covb dist=lnormal;  
run;
```

b) Détermination des primes stop-loss

```
# Les estimateurs ont été calculés à l'aide de SAS en a)
mu_2 <- 6.8418
sigma_2 <- 0.8278

# Valeurs des déductibles d à utiliser
d_2 <- 100 * seq(20,30)

# Calculer les primes STOP-LOSS pour les différents déductibles d
STOP_LOSS_2 <- function(x) {
  exp(mu_2 + sigma_2^2/2) * (1 - pnorm((log(x) - mu_2 - sigma_2^2)/sigma_2)) -
  x * (1 - pnorm((log(x) - mu_2)/sigma_2))
}

Primes_STOP_LOSS_2 <- data.frame(d_2, "Prime" = sapply(d_2, function(x) STOP_LOSS_2(x)))

# Présenter les valeurs des primes STOP-LOSS en fonction de d
library(xtable)
options(xtable.comment = FALSE)
xtable(Primes_STOP_LOSS_2, caption = "Valeurs de la prime STOP-LOSS pour
      chacun des déductibles d donnés",
      align = c("c", "c", "c"),
      digits = c(0, 0, 4))

# Représenter graphiquement l'évolution de la prime stop-loss en fonction de d
library(ggplot2)
plot5 <- ggplot(data = Primes_STOP_LOSS_2, aes(x= Primes_STOP_LOSS_2$d_2,
      y = Primes_STOP_LOSS_2$Prime))
  + geom_point() + geom_line() + xlab("d") + ylab("Prime stop-loss")
  + ggtitle("(A)") + theme(plot.title = element_text(face="bold", hjust = 0.5))

value_test_3 <- data.frame(d=(100 * seq(20,60)), "Prime" = sapply((100 * seq(20,60)),
      function(x) STOP_LOSS_2(x)))

plot6 <- ggplot(data=value_test_3, aes(x=value_test_3$d, y=value_test_3$Prime))
  + geom_point() + geom_line() + xlab("d") + ylab("Prime stop-loss") + ggtitle("(B)")
  + theme(plot.title = element_text(face="bold", hjust = 0.5))
```

c) Détermination des primes limited loss

```
# Valeurs des limites u à utiliser
u_2 <- 100 * seq(30,40)

# Calculer l'espérance limitée pour les différents limites u
Esp_limite_2 <- function(x) {
  exp(mu_2 + sigma_2^2/2) * pnorm((log(x) - mu_2 - sigma_2^2)/sigma_2) +
  x * (1 - pnorm((log(x) - mu_2)/sigma_2))
}

Primes_Esp_limite_2 <- data.frame(u_2, "Prime" = sapply(u_2, function(x) Esp_limite_2(x)))
```

```

# Présenter les valeurs des espérances limitées en fonction de u
library(xtable)
options(xtable.comment = FALSE)
xtable(Primes_Esp_limite_2, caption = "Valeurs de l'espérance limitée
    pour chacune des limites u données",
    align = c("c", "c", "c"),
    digits = c(0, 0, 4))

# Représenter graphiquement l'évolution de l'espérance limitée en fonction de u
library(ggplot2)
plot7 <- ggplot(data = Primes_Esp_limite_2, aes(x= Primes_Esp_limite_2$u_2,
    y = Primes_Esp_limite_2$Prime))
    + geom_point() + geom_line() + xlab("u") + ylab("Espérance limitée")
    + ggtitle("(C)") + theme(plot.title = element_text(face="bold", hjust = 0.5))

value_test_4 <- data.frame(u=(100 * seq(30,60)), "Prime" = sapply((100 * seq(30,60)),
    function(x) Esp_limite_2(x)))

plot8 <- ggplot(data=value_test_4, aes(x=value_test_4$u, y=value_test_4$Prime))
    + geom_point() + geom_line() + xlab("u") + ylab(" Espérance limitée")
    + ggtitle("(D)") + theme(plot.title = element_text(face="bold", hjust = 0.5))

# Mettre les 4 graphiques ensembles
library(gridExtra)
grid.arrange(plot5, plot6, plot7, plot8, ncol=2, nrow=2)

```

Question 3

a) Estimation du paramètre θ

Aucun code n'a été utilisé pour cette section.

b) Présentation graphique de la loi beta a priori

```

# Création du graphique de la loi beta
ggplot(data.frame(x=c(0, 1)),aes(x)) +
    stat_function(fun=dbeta, args = list(2,2))+
    xlab(expression(theta))+
    ylab("Densité")+
    geom_point(aes(x=0.5, y=1.5), colour="red")

```

c) Détermination de la loi a posteriori ainsi que la valeur exacte de l'estimateur de Bayes

Aucun code n'a été utilisé pour cette section.

d) Approximation de l'estimateur de Bayes par simulation

```
# Simulation de l'estimateur de Bayes  
set.seed(4545)  
nsim <- 500000  
simul_posteriorie <- rbeta(nsim, 4, 7)  
simul_bayes <- mean(simul_posteriorie)
```