

Mathématiques actuarielles IARD-1
ACT-2005
Notes de cours

Gabriel Crépeault-Cauchon
Nicholas Langevin

4 février 2019


Table des matières

Résumé

Ce document résume les notes de cours prises en classe dans le cours de Mathématiques actuarielles IARD-1, ainsi que des notions prises du livre *LOSS MODELS - From Data to Decisions, 4th edition*. Il est séparé en 2 parties : la matière qui a été couverte avant et après l'examen partiel.

Le chapitre correspondant dans le *Loss Models, 4th edition* est indiqué entre parenthèse des chapitres.

De plus, ce document essaie le plus possible de suivre la numérotation des sections du *Loss Models*. C'est pourquoi il arrive parfois qu'il y ait des sauts de section.

Un aide-mémoire a été créé en lien avec cette matière. Il est disponible [dans le répertoire GitHub suivant](#) .

Finalement, il y a une section à la fin des chapitres qui indique les exercices recommandés dans le livre, intitulée **Exercices recommandés**.

Première partie

Matière pour l'examen partiel

Chapitre 1

Quelques rappels de Probabilité

La loi gamma :

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

De plus, on a

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

et

$$\Gamma(n) = (n - 1)!$$

1.1 Fonction de survie

On a

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_{-\infty}^{\infty} S_X(x) dx \end{aligned}$$

De plus,

$$f_X(x) = -\frac{\partial}{\partial x} S_X(x)$$

Démonstration.

$$\begin{aligned}f_X(x) &= -\frac{\partial}{\partial x} S_X(x) \\&= -\frac{\partial}{\partial x} (1 - F_X(x)) \\&= -(-f_X(x)) \\&= f_X(x)\end{aligned}$$

□

Chapitre 2

Quantités des distributions à connaître (3)

2.1 Moments

Raw moments On représente le k^{e} moment par μ'_k , soit

$$\mu'_k = E \left[X^k \right] \quad (2.1)$$

Moments centraux Le k^{e} moment central est représenté par

$$\mu_k = E \left[(X - \mu)^k \right] \quad (2.2)$$

Exemple 2.1.1 Quelques exemples de moments centraux



La variance est le 2^e moment central :

$$\text{Var}(X) = \mu_2 = E \left[(X - \mu)^2 \right]$$

Le 3^e moment centré, qui est utilisé pour calculer le coefficient d'asymétrie :

$$\mu_3 = E \left[(X - \mu)^3 \right]$$

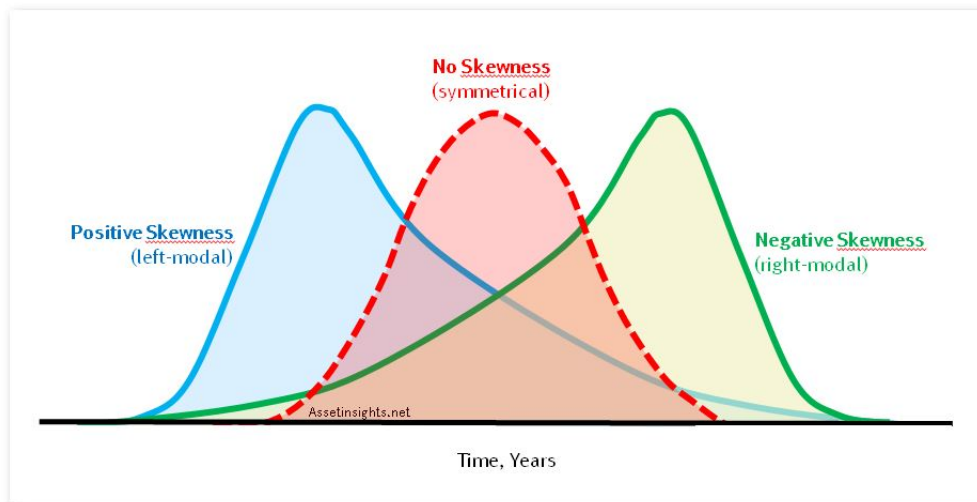
Coefficient de variation

$$CV = \frac{\sigma}{E[X]} \quad (2.3)$$

Coefficient d'asymétrie Le coefficient d'asymétrie, aussi appelé *skewness*, est défini par

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \quad (2.4)$$

Soit le 3^e moment standardisé. Si $S_k = 0$, alors la distribution tend vers une loi normale, telle qu'on le voit sur la figure ci-dessous :



Coefficient d'aplatissement Le coefficient d'aplatissement, aussi appelé *Kurtosis*, se définit par

$$\text{Kurtosis} = \frac{\mu_4}{\sigma^4} \quad (2.5)$$

Cette quantité permet de mesurer l'épaisseur de l'aile (*tail*) de la distribution. Si $E[z^4] = 3$, alors la distribution tend vers une loi normale $N(\mu, \sigma^2)$.

Mean Excess Loss On définit la variable aléatoire Y^P , qui représente le montant de perte en excès d'un déductible d , sachant que la perte est au

delà de ce montant. On peut définir l'espérance des coûts de cette variable aléatoire :

$$e(d) = E[Y^P] = E[X - d | X \geq d] = \frac{\int_d^\infty S_X(x)}{S_X(d)} \quad (2.6)$$

Note : cette variable est dite tronquée à gauche et *shifted*. On entend parfois aussi *Per-payment*

Left censored and shifted variable Soit la v.a. Y^L , qui représente le montant payé par l'assureur *par perte*. La variable est donc dite *censurée à gauche et shifted*. On peut aussi en calculer l'espérance :

$$E[Y^L] = E[(X - d)_+] = \int_d^\infty (x - d)f_X(x)dx \quad (2.7)$$

De plus, on peut facilement déduire la relation suivante :

$$E[(X - d)_+] = e(d)(1 - F_X(d))$$

Limited Loss Variable Finalement, on peut définir la variable Y , qui représente le paiement de l'assureur avec une limite de u à la police. Son espérance est définie par

$$E[X \wedge u] = \int_0^u f_X(x)dx \quad (2.8)$$

À l'aide d'intégration par partie, on peut trouver la forme suivante :

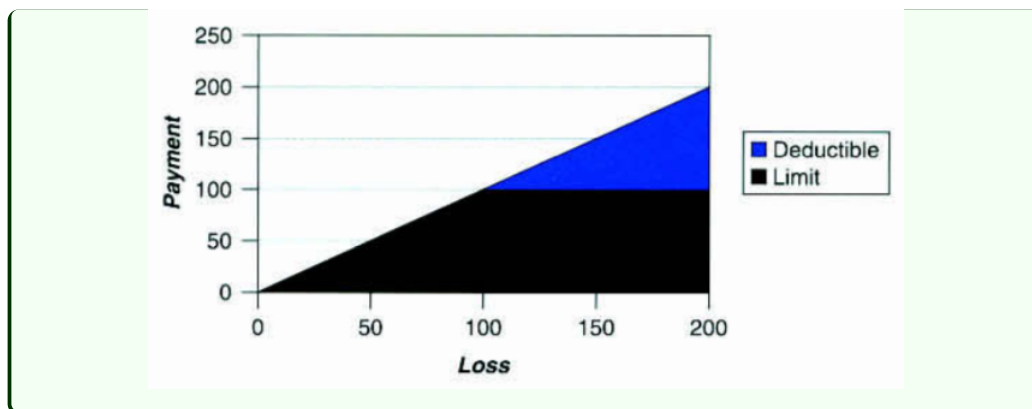
$$E[X \wedge u] = \int_0^u S_X(x)dx$$

La v.a. Y est dite *censurée à droite et shifted*

Exemple 2.1.2 lien entre le déductible et la limite



On peut faire le lien entre le déductible et la limite :



2.4 Percentile

On définit le p^e quantile (*percentile* en anglais) π_p comme étant la valeur minimale que X prend tel que $F_X(\pi_p) \geq p$.

Il arrive parfois (en contexte où x est discret) que le p percentile demandé tombe dans une *marche*. Il faut alors considérer les bornes de la fonction pour déterminer quelle valeur de x correspond au quantile demandé.

2.5 Queue de distribution

2.5.1 Classification selon les moments

- › On peut déterminer si une distribution a une *heavy-tail* en vérifiant si ses moments existent.
- › On peut aussi comparer des distributions entre-eux en utilisant des quantités standardisées, telles que le Coefficient de variation, le coefficient d'asymétrie (*skewness*) ou encore le coefficient d'aplatissement (*Kurtosis*)

2.5.2 Classification selon les comportements limites des ailes de distribution

› On peut faire le ratio de deux distributions avec leurs fonction de survie ($S(x)$) ou leur fonction f de densité pour vérifier laquelle des 2 a la plus grosse aile de distribution (*tail*).

1. Sois $f_1(x)$ une fonction tels que les 3 premiers moment existe : $E[x^4] = \infty$
2. Sois $f_2(x)$ une fonction tels que les 2 premiers moment existe : $E[x^2] = \infty$

Alors,

$$\lim_{x \rightarrow \infty} r(x) = \frac{f_1(x)}{f_2(x)} = \begin{cases} \infty, & f_1(x) \text{ a une aile plus lourde que } f_2(x) \\ 0, & f_2(x) \text{ a une aile plus lourde que } f_1(x) \end{cases}$$

Exemple 2.5.1

Sois $f_{x_1}(x_1) \sim \text{pareto}(\alpha, \theta)$ et $f_{x_2}(x_2) \sim \text{gamma}(\alpha, \lambda)$

$$\begin{aligned} \lim_{x \rightarrow \infty} r(x) &= \lim_{x \rightarrow \infty} \frac{f_{x_1}(x_1)}{f_{x_2}(x_2)} \\ &= \frac{\frac{\alpha \theta^\alpha}{(x+\theta)^{\alpha+1}}}{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}} \\ &= C \frac{e^{-\lambda x}}{x^{\alpha-1} (x+\theta)^{\alpha+1}} \\ &= \infty \end{aligned}$$

La pareto a une queue plus lourde que la gamma

2.5.3 Classification basée sur la fonction de Hazard

Définition 2.5.1 Hazard rate function



La fonction de hazard (aussi appelée force de mortalité ($\mu(x)$) ou le failure rate ($\lambda(x)$), est définie par

$$h_X(x) = \frac{f_X(x)}{S_X(x)} \quad (2.9)$$

On peut aussi exprimer la fonction $h_X(x)$ comme

$$h_X(x) = -\ln(S_X(x))$$

Soit une distribution ayant fonction de densité $f_X(x)$ et fonction de hazard $h_X(x)$. Alors,

- › Si $h(x) \nearrow$, *light-tailed*
- › Si $h(x) \searrow$, *heavy-tailed*
- › Note : on peut aussi comparer les distributions entre-elles : si une distribution voit son $h_1(x)$ augmenter plus rapidement que l'autre (i.e. $h_2(x)$), alors la deuxième distribution a une aile de distribution plus lourde.

2.6 Exercices recommandés

À compléter

Chapitre 3

Fréquence et sévérité avec modifications aux contrats (8)

3.2 Déductibles

3.2.1 Ordinary deductible

Définition

Soit un contrat d'assurance avec déductible d . Lors d'une perte, l'assureur va payer tout montant en excédent du montant d . Alors, pour la variable *per-payment*,

$$Y^L = (X - d)_+ = \begin{cases} 0 & , X \leq d \\ X - d & , X > d \end{cases} \quad (3.1)$$

$$Y^P = (X - d)_+ = \begin{cases} \text{Non-défini} & , X \leq d \\ X - d & , X > d \end{cases} \quad (3.2)$$

Fonctions reliées

Et on peut aussi déduire toutes les fonctions qui y sont reliées :

$$f_{Y^P}(y) = \frac{f_X(y + d)}{S_X(d)}$$

$$S_{Y^P}(y) = \frac{S_X(y + d)}{S_X(d)}$$

$$F_{Y^P}(y) = \frac{F_X(y + d) - F_X(d)}{S_X(d)}$$

$$h_{Y^P}(y) = h_X(y + d)$$

Pour la v.a. Y^L *per-loss*¹,

$$f_{Y^L}(y) = f_X(y + d)$$

$$S_{Y^L}(y) = S_X(y + d)$$

$$F_{Y^L}(y) = F_X(y + d)$$

Espérance

$$\mathbb{E} [Y^L] = \mathbb{E} [(X - d)_+] = \mathbb{E} [X] - \mathbb{E} [X \wedge d] \quad (3.3)$$

$$\mathbb{E} [Y^P] = \frac{\mathbb{E} [(X - d)_+]}{S_X(d)} = \frac{\mathbb{E} [X] - \mathbb{E} [X \wedge d]}{S_X(d)} \quad (3.4)$$

1. Il est à noter que la fonction de Hazard n'est pas définie à 0.

Cette espérance s'appelle la prime *Stop-Loss*, et elle est définie par

$$\begin{aligned}
 E[Y] &= E[(X - d)_+] \\
 &= \int_d^\infty (x - d)f(x)dx \\
 &= \underbrace{\int_d^\infty xf(x)dx}_{\text{Intégration par partie}} - d \int_d^\infty f(x)dx \\
 &= -xS(x) \Big|_d^\infty + \int_d^\infty S(x)dx - dS(d) \\
 &= 0 + S(d) + \int_d^\infty S(x)dx - S(d) \\
 &= \int_d^\infty S(x)dx
 \end{aligned}$$

3.2.2 Franchise deductible

Définitions

Lorsque la perte dépasse le deductible franchise de montant d , l'assureur assume l'entièreté des coûts². Pour la v.a. Y^L , on a

$$Y^L = \begin{cases} 0 & , X \leq d \\ X & , X > d \end{cases}$$

Pour la v.a. Y^P ,

$$Y^P = \begin{cases} \text{non-défini} & X \leq d \\ X & , X > d \end{cases} \quad (3.5)$$

2. On voit plus souvent ce type de deductible dans un contexte d'invalidité : si on est absent plus d'un certain nombre de jours du travail, on se fait rembourser toutes ses absences en salaire.

Fonctions reliées

Les fonctions de la v.a. Y^L sont

$$f_{Y^L}(y) = \begin{cases} F_X(d) & , y = 0 \\ f_X(y) & , y > 0 \end{cases}$$

$$F_{Y^L}(y) = \begin{cases} F_X(d) & , 0 < y \leq d \\ F_X(y) & , y \geq d \end{cases}$$

$$S_{Y^L}(y) = \begin{cases} S_X(d) & , 0 < y \leq d \\ S_X(x) & , y > d \end{cases}$$

$$h_{Y^L}(y) = \begin{cases} h_X(d) & , 0 < y \leq d \\ h_X(x) & , y > d \end{cases}$$

Pour la fonction Y^P (*per-payment*),

$$f_{Y^P}(y) = \frac{f_X(d)}{S_X(d)}$$

$$F_{Y^P}(y) = \begin{cases} F_X(d) & , y = 0 \\ F_X(d) & , 0 < y \leq d \\ \frac{F_X(y) - F_X(d)}{S_X(d)} & , y > d \end{cases}$$

$$S_{Y^P}(y) = \begin{cases} 1 & , 0 \leq y \leq d \\ \frac{S_X(y)}{S_X(d)} & , y > d \end{cases}$$

$$h_{Y^P}(y) = \begin{cases} 0 & , 0 < y \leq d \\ h_X(y) & , y > d \end{cases}$$

Espérance

$$E[Y^L] = E[(X - d)_+^F] = E[X] - E[X \wedge d] + dS_X(d) \quad (3.6)$$

$$E[Y^P] = E[(X - d)_+^F] = \frac{E[X] - E[X \wedge d]}{S_X(d)} + d \quad (3.7)$$

3.3 Loss Elimination Ratio

Définition 3.3.1 Loss Eliminating Ratio



Le Loss Eliminating Ratio (*LER*), nous permet d'obtenir le pourcentage de perte qu'on ne paiera pas grâce au déductible d :

$$LER = \frac{E[X] - E[(X - d)_+]}{E[X]}$$

Mais on sait que

$$E[(X - d)_+] = E[X] - E[X \wedge d]$$

Alors,

$$LER = \frac{E[X \wedge d]}{E[X]} \quad (3.8)$$

Note sur l'inflation Il arrive en exercice qu'on nous demande de comparer ce ratio avec et sans inflation. Soit un contrat avec r % d'inflation. Alors, on peut trouver $E[(X - d)_+]$ qui tient compte de cette inflation :

$$E[(X - d)_+^I] = (1 + r) \left(E[X] - E\left[X \wedge \frac{d}{1+r}\right] \right)$$

Démonstration.

$$E[Y] = (1 + r) \int_{\frac{d}{1+r}}^{\infty} x f_X(x) dx - d \int_{\frac{d}{1+r}}^{\infty} f_X(x) dx$$

En ajoutant un terme,

$$\begin{aligned}
&= \underbrace{(1+r) \int_0^{\frac{d}{1+r}} x f_X(x) dx + \int_{\frac{d}{1+r}}^{\infty} x f_X(x) dx}_{E[X]} - \underbrace{\int_0^{\frac{d}{1+r}} x f_X(x) dx - \frac{d}{1+r} \int_{\frac{d}{1+r}}^{\infty} f_X(x) dx}_{E\left[X \wedge \frac{d}{1+r}\right]} \\
&= E[X] - E\left[X \wedge \frac{d}{1+r}\right]
\end{aligned}$$

□

3.4 Policy Limits

3.4.1 Définitions

Soit un contrat d'assurance où il est conclu que l'assureur débourse un maximum de u dans le montant de la perte. Ce type de modification au contrat crée une v.a. *censurée à droite*, i.e. le montant déboursé est maximisé à u . On définit Y comme étant

$$Y = \begin{cases} X & , X \leq u \\ u & , X > u \end{cases} \quad (3.9)$$

3.4.2 Fonctions reliées

On peut déduire les fonctions reliées suivantes :

$$f_Y(y) = \begin{cases} f_X(y) & , y < u \\ S_X(u) & , y = u \end{cases}$$

$$F_Y(y) = \begin{cases} F_X(y) & , y \leq u \\ 1 & , y > u \end{cases}$$

Note sur l'inflation Si on a r % d'inflation, alors

$$E[X \wedge u] = (1+r)E\left[X \wedge \frac{u}{1+r}\right]$$

Démonstration.

$$\begin{aligned} E[Y] &= \int_0^{\frac{u}{1+r}} (1+r)x f_X(x) dx + u \int_{\frac{u}{1+r}}^{\infty} f_X(x) dx \\ &= (1+r) \int_0^{\frac{d}{1+r}} x f_X(x) dx + S_X\left(\frac{u}{1+r}\right) \\ &= (1+r) \left(\int_0^{\frac{u}{1+r}} x f_X(x) dx + \frac{u}{1+r} S_X\left(\frac{u}{1+r}\right) \right) \\ &= (1+r)E\left[X \wedge \frac{u}{1+r}\right] \end{aligned}$$

□

3.5 Coassurance, déductible et limites

Définitions

Le livre nous donne une fonction de perte qui englobe 4 éléments en même temps : l'inflation, les déductibles, la coassurance³ et les limites de police. Pour la v.a. Y^L ,

$$Y^L = \begin{cases} 0 & , x < \frac{d}{1+r} \\ \alpha((1+r)x - d) & , \frac{d}{1+r} \leq x < \frac{u}{1+r} \\ \alpha(u - d) & , x \geq \frac{u}{1+r} \end{cases} \quad (3.10)$$

et pour Y^P ,

$$Y^L = \begin{cases} \text{Non-défini} & , x < \frac{d}{1+r} \\ \alpha((1+r)x - d) & , \frac{d}{1+r} \leq x < \frac{u}{1+r} \\ \alpha(u - d) & , x \geq \frac{u}{1+r} \end{cases} \quad (3.11)$$

3. Dans ce type de contrat, la compagnie paie une proportion α de la perte, et le $(1 - \alpha)$ restant est assumé par le titulaire de la police.

Espérance

$$\mathbb{E} \left[Y^L \right] = \alpha(1+r) \left(\mathbb{E} \left[X \wedge \frac{u}{1+r} \right] - \mathbb{E} \left[X \wedge \frac{d}{1+r} \right] \right) \quad (3.12)$$

$$\mathbb{E} \left[Y^P \right] = \frac{\mathbb{E} \left[Y^L \right]}{S_X \left(\frac{d}{1+r} \right)} \quad (3.13)$$

$$\begin{aligned} \mathbb{E} [Y] &= \int_{\frac{d}{1+r}}^{\frac{u}{1+r}} ((1+r)x - d) f_X(x) dx + \int_{\frac{u}{1+r}}^{\infty} (u - d) f_X(x) dx \\ &= (1+r) \left(\mathbb{E} \left[X \wedge \frac{u}{1+r} \right] - \mathbb{E} \left[X \wedge \frac{d}{1+r} \right] \right) \end{aligned}$$

3.6 Exercices recommandés

À compléter

Chapitre 4

Estimation de données complètes (11)

4.2 Estimation de données complètes

4.2.1 Estimation de la fonction de répartition empirique

On cherche à estimer $F(t)$ ou $S(t)$, lorsque nos données sont complètes (i.e. x_1, \dots, x_n qui sont *iid*). Alors, l'estimateur non paramétrique pour $F(t)$:

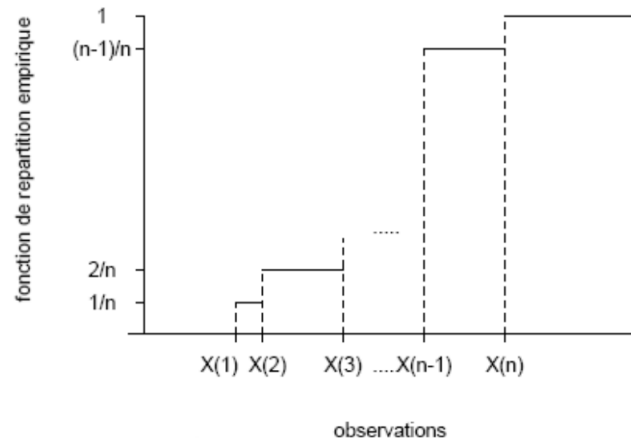
$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i \leq t\}} \quad (4.1)$$

où $\mathbb{1}_{\{.\}}$ représente une fonction indicatrice.

$F_x(t)$ aura donc la forme suivante :

$$F_n(t) = \begin{cases} 0 & , t < x_{(1)} \\ \frac{1}{n} & , x_{(1)} \leq t < x_{(2)} \\ \frac{2}{n} & , x_{(2)} \leq t < x_{(3)} \\ \dots & \\ 1 & , t \geq x_{(n)} \end{cases} \quad (4.2)$$

où $t \in [0, x_{(n)}]$.



Remarques

(1) Lorsque $F_n(t) \rightarrow F(t)$, alors

(2) Puisqu'on a

$$\sum_{i=1}^n 1_{[x_i \leq t]} \sim \text{Bin}(n, \Pr(X \leq t))$$

Alors,

$$\begin{aligned} E[F_n(t)] &= \frac{1}{n} n F(t) \\ &= F(t) \quad (\text{C'est un estimateur sans biais}) \\ \text{Var}(F_n(t)) &= \frac{1}{n^2} n F(t) S(t) \\ &= \frac{F(t) S(t)}{n} \\ &\underset{n \rightarrow \infty}{=} 0 \end{aligned}$$

4.2.2 Cumulative hazard-rate function

On utilise cette fonction pour pouvoir réussir à estimer la fonction de densité et le *hazard rate* empirique. En effet, puisque la distribution empirique

est discrète, on ne peut pas dériver $F_n(x)$ pour obtenir $f_n(x)$ et $h_n(x)$. La fonction de hazard cumulative se définit par

$$H_X(x) = -\ln S_X(x) \quad (4.3)$$

4.2.3 Notation à utiliser pour la distribution empirique

- › $y_1 < y_2 < \dots < y_k$: les k valeurs uniques qui apparaissent dans l'échantillon de n . **Note** : $k \leq n$.
- › s_j : nombre de fois que l'observation y_j est observée dans l'échantillon. On a

$$\sum_{j=1}^k s_j = 1$$

- › r_j : nombre d'observations qui sont plus grandes ou égales à y_j . On a

$$r_j = \sum_{i=j}^k s_i$$

- › Avec cette nouvelle notation, on peut ré-écrire la fonction de répartition empirique :

$$F_n(x) = \begin{cases} 0 & , x < y_1 \\ 1 - \frac{r_j}{n} & , y_{j-1} \leq x \leq y_j \quad j = 2, \dots, k \\ 1 & , x \geq y_k \end{cases} \quad (4.4)$$

4.2.4 Estimateur de Nelson Åalen

Pour pouvoir estimer la *Cumulative hazard-rate function*, on doit utiliser un estimateur qui se base sur la notation utilisée à la sous-section précédente, soit le *Nelson Åalen estimate* :

$$\hat{H}(x) = \begin{cases} 0 & , x < y_1 \\ \sum_{i=1}^{j-1} \frac{s_i}{r_i} & , y_{j-1} \leq x \leq y_j \quad j = 2, \dots, k \\ \sum_{i=1}^k \frac{s_i}{r_i} & , x \geq y_k \end{cases} \quad (4.5)$$

4.3 Distribution empirique avec données groupées

4.3.1 Fonction OGIVE

CETTE MATIÈRE NE SERA PAS TESTÉE À L'EXAMEN intra A2018

Dans certains contextes, on a n données qui sont groupées en intervalles. La fonction OGIVE permet d'interpoler entre 2 points c_{j-1} et c_j .

$$\begin{aligned} c_{j-1} &\leq x \leq c_j \\ F_n(c_{j-1}) &\leq F_n(x) \leq F_n(c_j) \end{aligned}$$

On peut déterminer la valeur empirique de $F_n(x)$ aux bornes des classes, avec

$$F_n(x) = \frac{\sum_{i=1}^j n_i}{n}$$

où n_i est le nombre d'observations qui sont entre c_{j-1} et c_j . Toutefois, pour les valeurs entre deux bornes $c_0 < c_1 < \dots < c_k$, il faut approximer avec la fonction OGIVE ci-dessous :

$$F_n^{\text{OGIVE}}(x) = \frac{c_j - x}{c_j - c_{j-1}} F_n(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}} F_n(c_j) \quad (4.6)$$

Remarques

(1) Si $x = c_{j-1}$,

$$F_n(c_{j-1}) = F_n^{\text{OGIVE}}(c_{j-1})$$

(2) Si $x = c_j$,

$$F_n(c_j) = F_n^{\text{OGIVE}}(c_j)$$

(3) En dérivant (??), on obtient

$$\begin{aligned} f_n(x) &= \frac{\partial}{\partial x} F_x(x) \\ &= \frac{1}{c_j - c_{j-1}} F_x(c_j) - \frac{1}{c_j - c_{j-1}} F_n(c_{j-1}) \\ &= \frac{F_n(c_j) - F_n(c_{j-1})}{c_j - c_{j-1}} \end{aligned} \tag{4.7}$$

Exemple 4.3.1 Fonction R ogive



Du package [actuar](#), on peut utiliser la fonction `ogive()` pour calculer les probabilités empiriques à des points précis de nos données lorsqu'elles sont groupées, mais aussi interpoler entre 2 valeurs :

```
## Fonction OGIVE du package actuar de Vincent Goulet
## Exemple numéro 5.2 du Loss Models
## page 68, 4e édition
library(actuar)

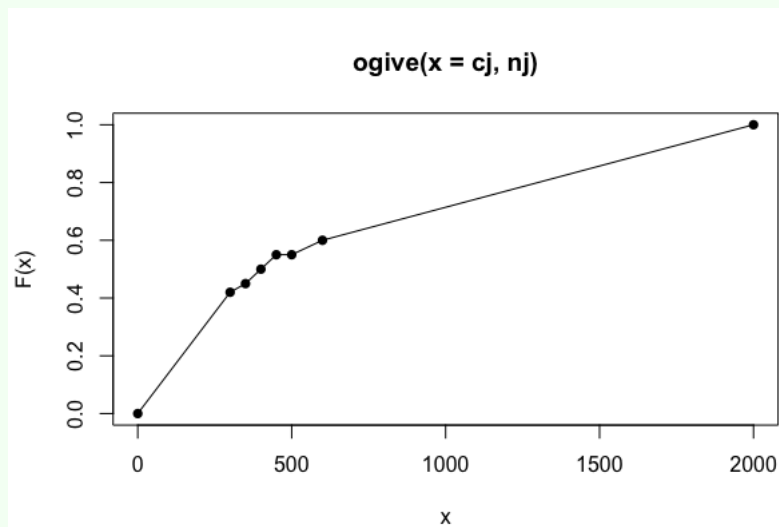
## On définit le vecteur cj des bornes et
## un vecteur nj des fréquences empiriques
# note : il faut forcer une borne supérieure
cj <- c(0, 300, 350, 400, 450, 500, 600, 2000)
nj <- c(42, 3, 5, 5, 0, 5, 40)

# On calcule la fonction ogive avec la fonction du même nom
Fn <- ogive(cj, nj)

y <- 500 / (1.1)^3
# On calcule la probabilité demandée
1 - Fn(y)
[1] 0.5243426

# On peut aussi faire un graphique de la
## fonction empirique OGIVE
plot(Fn)
```

Et on obtient le graphique suivant :



4.4 Exercices recommandés

À compléter

Chapitre 5

Estimation de données modifiées (12)

5.1 Point estimation

5.1.1 Définitions importante

Un vocabulaire spécifique aux données modifiées est utilisé, soit

Définition	Terme anglais	Explication
Tronquée à gauche	<i>left truncated at d</i>	si la valeur observée est plus basse que d , elle n'est pas enregistrée
Tronquée à droite	<i>right truncated at u</i>	si la valeur observée est plus grande que u , elle n'est pas enregistrée
Censurée à gauche	<i>left censored at d</i>	si la valeur observée est plus basse que d , on indique d dans les données modifiées
Censurée à droite	<i>right censored at u</i>	si la valeur observée est plus grande que u , on indique u dans les données modifiées

Note : Il arrive aussi que les données soient *shifted at d* , ce qui veut dire qu'on soustrait d aux données (souvent en présence d'un déductible).

Notation

- > $y_1 < y_2 < \dots < y_k$: les k valeurs uniques qui apparaissent dans l'échantillon.
- > x_j : la j^{e} données non-censurée qui apparait dans l'échantillon
- > d_j : le montant auquel x_j est tronquée. Si il n'y a pas de troncage, alors $d_j = 0$.
- > u_j : la j^{e} données censurée qui apparait dans l'échantillon.
- > r_j : nombre d'observations qui sont plus grandes ou égales à y_j (*risk*

set)

$$r_j = \# \text{ of } x_i + \# \text{ of } u_i - (\# \text{ of } d_i \geq y_j)$$

> s_i : nombre de décès au temps i .

Estimateur de Kaplan-Meier

Cet estimateur est une version modifiée de l'estimateur de Nelson-Åalen vu à la section ?? . Avec l'information des données et en utilisant la notation vue à la sous-section précédente, on peut estimer la fonction de survie empirique :

$$\hat{S}_n(t) = \begin{cases} 1 & , 0 \leq t < y_1 \\ \prod_{i=1}^{j-1} \left(\frac{r_i - s_i}{r_i} \right) & , y_{j-1} \leq t < y_j \quad j = 2, \dots, k \\ \prod_{i=1}^k \left(\frac{r_i - s_i}{r_i} \right) & t \geq y_k \end{cases} \quad (5.1)$$

5.2 Espérance, Variance et et intervalle d'estimation

Contexte

On s'intéresse à l'espérance et la variance de la fonction de survie S_n , qui suite une binomiale (i.e. $S_n(x) \sim \text{Bin}(n, S(x))$). Si on connaissait $S(x)$, on pourrait facilement déduire l'espérance et la variance avec les formules qu'on connaît. Toutefois, puisqu'on cherche souvent à estimer $S(x)$, il faudra aussi estimer l'espérance et la variance.

Section à compléter

5.3 Exercices recommandés

À compléter

Deuxième partie

Matière pour l'examen final

Chapitre 6

Continuous models (5)

6.1 Multiplication par une constante

Soit X la v.a. représentant les pertes et θ un paramètre d'échelle (*scale parameter*). On peut définir la v.a. Y par

$$Y = \theta X$$

Alors,

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) \\ &= \Pr(\theta X \leq y) \\ &= \Pr\left(X \leq \frac{y}{\theta}\right) \\ &= F_X\left(\frac{y}{\theta}\right) \end{aligned} \tag{6.1}$$

Et on peut aussi trouver que

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\ &= \frac{\partial}{\partial y} \left\{ F_X\left(\frac{y}{\theta}\right) \right\} \\ &= \frac{1}{\theta} f_X\left(\frac{y}{\theta}\right) \end{aligned} \tag{6.2}$$

À noter Si on doit appliquer une transformation impliquant une puissance (*raising to a power*, voir section ??), on applique cette dernière transformation avant d'appliquer le paramètre d'échelle θ .

6.2 Raising to a power

Soit X une v.a. représentant la perte avec une loi de probabilité quelconque et la v.a. Y tel que

$$Y = X^{\frac{1}{\tau}}$$

avec $\tau > 0$. On peut trouver $F_Y(y)$ ainsi que $f_Y(y)$:

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) \\ &= \Pr\left(X^{\frac{1}{\tau}} \leq y\right) \\ &= \Pr(X \leq y^\tau) \\ &= F_X(y^\tau) \end{aligned} \tag{6.3}$$

et

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\ &= \frac{\partial}{\partial y} F_X(y^\tau) \tau y^{\tau-1} f_X(y^\tau) \end{aligned} \tag{6.4}$$

Note On préfère toujours des paramètres positifs, donc on va ajuster les formules avec des constantes négatives dans certains cas.

Exemple 6.2.1 Notation dans le livre *Loss Models*



La notation utilisée par le livre est la suivante :

- > si $\tau > 0$: *transformed*
- > si $\tau = -1$: *inverse*
- > si $\tau < 0$ et $\tau \neq -1$: *inverse-transformed*

6.3 Exponentiation

Soit X une v.a. représentant la perte. On définit Y comme

$$Y = \exp(X)$$

Alors, on trouve

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) \\ &= \Pr(e^X \leq y) \\ &= \Pr(X \leq \ln y) \\ &= F_X(\ln y) \end{aligned} \tag{6.5}$$

et

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\ &= \frac{\partial}{\partial y} \{F_X(\ln y)\} \\ &= \frac{1}{y} f_X(\ln y) \end{aligned} \tag{6.6}$$

Note La loi Log-Normale est en fait une transformation exponentielle de la Loi Normale.

6.4 Loix mélanges

On peut créer des loix mélanges. En effet, il arrive qu'on ait une distribution dont l'un des paramètres est lui-même distribué aléatoirement selon une autre distribution, tel que

$$X|\Lambda \sim \text{une certaine loi}$$

$$\Lambda \sim \text{une autre loi (ou la même!)}$$

On peut trouver les fonctions de densité et de répartition de ces différentes lois :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|\Lambda}(x|\lambda) f_{\Lambda}(\lambda) d\lambda \quad (6.7)$$

$$F_X(x) = \int_{-\infty}^{\infty} F_{X|\Lambda}(x|\lambda) f_{\Lambda}(\lambda) d\lambda \quad (6.8)$$

De plus, on peut trouver l'espérance et la variance de ces lois mélanges avec les théorèmes de l'espérance et la variance conditionnelle :

$$E[X] = E[E[X|\Lambda]] \quad (6.9)$$

$$Var(X) = E[Var(X|\Lambda)] + Var(E[X|\Lambda]) \quad (6.10)$$

6.5 Frailty models

Est sur le syllabus des examens professionnels, mais ne semble pas être à l'étude pour le cours IARD1 ...

Soit un paramètre $\Lambda > 0$ (*frailty random variable*) qu'on va utiliser pour trouver la *hazard-rate* conditionnel $h_{X|\Lambda}$ tel que

$$h_{X|\Lambda}(x|\lambda) = \lambda a(x) \quad (6.11)$$

où $a(x)$ est une fonction quelconque et la distribution de $\Lambda = \lambda$ est connu. Alors, on peut trouver la fonction de survie de $S_{X|\Lambda}(x|\lambda)$ avec des des propriétés que l'on connaît :

$$\begin{aligned} S_{X|\Lambda}(x|\lambda) &= e^{-\int_0^x h_{X|\Lambda}(t|\lambda) dt} \\ &= e^{-\lambda A(x)} \\ &= M_{\Lambda}(-A(x)) \\ &= \mathcal{L}_{\Lambda}(A(x)) \end{aligned}$$

6.6 Splicing

Soit X_1, \dots, X_k les différentes distributions que peut prendre la v.a. X et α_i la probabilité que $X \sim X_i$, $i = 1, 2, \dots, k$. Pour $\sum \alpha_i = 1$, on a donc

$$f_X(x) = \begin{cases} \alpha_1 f_1(x) & , c_0 < x < c_1 \\ \alpha_2 f_2(x) & , c_1 < x < c_2 \\ \dots & \\ \alpha_k f_k(x) & , c_{k-1} < x < c_k \end{cases}$$

Note Dans un problème, on peut soit demander de calculer la fonction de répartition à un certain point, ou encore nous demander de trouver les coefficients α_i en connaissant les différentes fonctions de répartition et les probabilités rattachées.

6.7 Exercices recommandés

Les exercices recommandés dans le *Loss Models, 4th edition* pour ce chapitre sont les suivants :

- > 5.1 à 5.6
- > 5.8 à 5.11
- > 5.17 à 5.19

Chapitre 7

Modèle de perte agrégée (9)

7.1 Modèle composé (fréquence-sévérité) pour les pertes agrégées

On définit les variables aléatoires suivantes :

- › S : Perte totale pour le portefeuille au complet;
- › N : Nombre de réclamations observées dans l'intervalle de temps. (**fréquence**);
- › X_1, \dots, X_n : Montant de perte relié à la i^e réclamation (**sévérité**).¹

On a donc

$$S = X_1 + \dots + X_n$$

et on peut trouver diverses quantités, tel que la fonction de répartition :

$$F_S(x) = \sum_{n=0}^{\infty} F_X^{*n}(x) \Pr(N = n) \quad (7.1)$$

où $F_X^{*n} = \Pr(X_1 + \dots + X_n \leq x)$.

On peut aussi obtenir la fonction génératrice des moments :

$$\mathcal{M}_S(t) = \mathcal{P}_N(\mathcal{M}_X(t)) \quad (7.2)$$

1. Les X_i sont iid et indépendants de N .

L'espérance et la variance peuvent être obtenues avec les propriétés d'espérance et variance conditionnelle :

$$\begin{aligned}
 E[S] &= E[E[S|N]] \\
 &= E\left[E\left[\sum_{i=1}^n X_i\right]\right] \\
 &= E\left[\sum_{i=1}^n E[X_i]\right] \\
 &\stackrel{iid}{=} E[NE[X]] \\
 &= E[N]E[X]
 \end{aligned}
 \tag{7.3}$$

$$\begin{aligned}
 \text{Var}(S) &= \text{Var}(E[S|N]) + E[\text{Var}(S|N)] \\
 &= \text{Var}\left(E\left[\sum_{i=1}^n X_i\right]\right) + E\left[\text{Var}\left(\sum_{i=1}^n X_i\right)\right] \\
 &= \text{Var}\left(\sum_{i=1}^n E[X_i]\right) + E\left[\sum_{i=1}^n \text{Var}(X_i)\right], \text{ par indépendance des } X_i \\
 &\stackrel{iid}{=} \text{Var}(NE[X]) + E[N\text{Var}(X)] \\
 &= E[X]^2 \text{Var}(N) + E[N] \text{Var}(X)
 \end{aligned}
 \tag{7.4}$$

7.2 Exercices recommandés

Les exercices recommandés dans le *Loss Models, 4th edition* pour ce chapitre sont les suivants :

› 9.34 et 9.36

Chapitre 8

Frequentist estimation (13)

8.1 Méthode des moments

On cherche à estimer les paramètres de la distribution de nos données, sachant que la distribution suit une loi quelconque.

Si on a p paramètres à estimer, alors il faudra calculer les p premiers moments empiriques de la distribution, soit

$$\hat{\mu}'_k = \sum_{i=1}^n x_i^k \quad (8.1)$$

Puis d'utiliser cette quantité empirique pour résoudre le système d'équation à p inconnus tel que $\hat{\mu}'_k = \mu'_k$.

8.2 Méthode des percentiles

Pour un rappel sur les percentiles, voir la ???. Cette méthode est appelée en anglais *percentile matching estimate*.

Un peu comme la méthode des moments, on utilise une quantité qu'on peut calculer empiriquement, soit la fonction de répartition empirique, pour pouvoir résoudre le système d'équations à p variables tel que

$$F_n(\hat{\pi}_{g_i}) = g_i$$

où

- › $\hat{\pi}_{g_i} = F_n^{-1}(g_i), i = 1, \dots, p$
- › F_n : la fonction de répartition empirique¹
- › p : nombre de paramètres de la distribution à estimer.

8.2.1 Smoothed empirical estimate

Il arrive parfois que le percentile arrive entre 2 *marche* de la fonction de répartition empirique. Dans cette situation, $\hat{\pi}_g$ doit être approximé de façon linéaire avec les statistiques d'ordre :

$$\hat{\pi}_g = (1 - h)X_{(j)} + hX_{(j+1)} \quad (8.2)$$

où

- › $j = \lfloor (n + 1)g \rfloor$ et $h = (n + 1)g - j$
- › $X_{(j)}$ la j^{e} statistique d'ordre de l'échantillon empirique X .

8.3 Méthode du maximum de vraisemblance

8.3.1 Données complètes

aussi appelé *Maximum log-likelihood estimator (MLE)*. On cherche à estimer le paramètre θ en maximisant la fonction de vraisemblance².

Définition 8.3.1 Fonction de vraisemblance



La fonction de vraisemblance $L(\theta|x)$ est la probabilité d'obtenir un échantillon x_1, \dots, x_n d'une distribution $f_X(x)$ ayant le vecteur de paramètres θ . Alors,

$$L(\theta) = \prod_{i=1}^n f_X(x_i; \theta) \quad (8.3)$$

1. L'estimation de la fonction de répartition empirique est couvert au ??.
2. Il arrive parfois que la seule façon de trouver le paramètre qui maximise $L(\theta)$ est en utilisant des [méthodes numériques](#)

On peut aussi trouver la fonction de log-vraisemblance $\ell(\theta)$ (qui est souvent plus facile à dériver) :

$$\ell(\theta) = \sum_{i=1}^n \ln f_X(x_i; \theta) \quad (8.4)$$

Estimation du paramètre θ On détermine l'estimateur du maximum de vraisemblance (*MLE* en anglais)³ tel qu'on maximise l'?. Intuitivement, on va donc trouver le paramètre θ tel que

$$\frac{\partial}{\partial \theta} L(\theta) = 0$$

Puisque la fonction de vraisemblance peut parfois être difficile à dériver, on utilise souvent la fonction de log-vraisemblance (??) :

$$\frac{\partial}{\partial \theta} \ell(\theta) = 0$$

Estimation du vecteur de paramètres θ Si on doit estimer plusieurs paramètres θ_i pour la fonction f_X , on procède de la même façon, en dérivant partiellement la fonction de vraisemblance (ou log-vraisemblance) par rapport à chacun des paramètres θ_i et en posant ces dérivées égales à 0.

8.3.2 Données groupées

Lorsque les données sont groupées dans les intervalles $c_0 < c_1 < \dots < c_k$, la fonction de vraisemblance $L(\theta)$ est différente. $X \sim \text{Multinomiale}(\mathbf{n}, \mathbf{p})$. Alors,

$$L(\theta) = \Pr(n_1, n_2, \dots, n_k; \theta) = \binom{n}{n_1, \dots, n_k} p_1^{n_1} \dots p_k^{n_k} \quad (8.5)$$

Dans le contexte de données groupées, on a que $p_j = \Pr(X_j \in (c_{j-1}, c_j))$. Donc, la fonction de vraisemblance est

$$L(\theta) = \prod_{j=1}^n (F_X(c_j | \theta) - F_X(c_{j-1} | \theta))^{n_j}$$

3. L'abréviation en français est EMV, mais on utilise surtout *MLE* pour Maximum likelihood estimator.

Et la fonction de log-vraisemblance (en enlevant la constante combinatoire) est donc

$$\ell(\theta) = \sum_{j=1}^k n_j \ln (F_X(c_j; \theta) - F_X(c_{j-1}; \theta)) \quad (8.6)$$

8.3.3 Données censurées

Si une donnée est censurée, on l'interprète comme une donnée groupée dont l'intervalle va jusqu'à ∞ . On remplace $f_X(x_j)$ par $1 - F_X(x_j)$.

8.3.4 Données tronquées

Si une donnée est tronquée, on a 2 options :

1. On soustrait de chacune des observations le point de troncation d ;
2. **(Méthode plus souvent utilisée)** : On conditionne pour que les données sous le point de troncation d ne soient pas considérées, i.e. on remplace $f(x_j|\theta)$ par $\frac{f(x_j|\theta)}{1-F(d)}$.

8.4 Calcul de la variance et intervalle de confiance

L'explication donnée ici est reprise de l'ouvrage *Modélisation des sinistres avec R*⁴.

Lorsque $n \rightarrow \infty$, on sait que le vecteur de paramètres $\hat{\theta} \sim N(\theta, \text{Var}(\hat{\theta}))$, avec $\text{Var}(\hat{\theta}) = I(\theta)^{-1}$ ⁵.

Définition 8.4.1 Information de Fisher



La matrice d'information de Fisher est définie par

$$I(\theta) = -nE \left[\frac{\partial^2}{\partial \theta^2} \ln f_X(x; \theta) \right] = nE \left[\left(\frac{\partial}{\partial \theta} \ln f_X(x; \theta) \right)^2 \right] \quad (8.7)$$

La première version étant seulement applicable dans le cas où toutes

4. [Ouvrage de Vincent Goulet](#), section 5.5 (page 203 à 212)

5. Pour un rappel sur les matrices inverses, voir l'annexe ??.

les variables et observations sont *iid*, on utilise plus souvent la version généralisée :

$$I(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = E \left[\left(\frac{\partial}{\partial \theta} \ell(\theta) \right)^2 \right] \quad (8.8)$$

Il arrive qu'on remplace l'espérance par *l'observation observée* dans les équations ci-dessus, lorsqu'elle est trop compliquée à calculer. On définit *l'information observée* telle que

$$\hat{I}(\hat{\theta}) = \sum_{i=1}^n \left(\frac{\partial}{\partial \theta} \ln f(x_i; \theta) \Big|_{\theta=\hat{\theta}} \right)^2 = - \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \ln f(x_i; \theta) \Big|_{\theta=\hat{\theta}} \quad (8.9)$$

Donc, on peut calculer un intervalle de confiance (bilatéral) au seuil $1 - \alpha$ autour de θ :

$$\Pr \left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sqrt{\text{Var}(\hat{\theta})}} < z_{\alpha/2} \right) = 1 - \alpha$$

$$\Pr \left(\hat{\theta} - z_{\alpha/2} \sqrt{\text{Var}(\hat{\theta})}, \hat{\theta} + z_{\alpha/2} \sqrt{\text{Var}(\hat{\theta})} \right) = 1 - \alpha$$

D'où l'IC au seuil $1 - \alpha$ est

$$\theta \in \left[\hat{\theta} \pm z_{\alpha/2} \sqrt{\text{Var}(\hat{\theta})} \right] \quad (8.10)$$

8.4.1 Méthode delta

Dans certains cas, on désire approximer une autre quantité que le paramètre lui-même, qui est en fait une transformation $h(\theta)$ de ce dernier. On doit utiliser la méthode Delta, qui fait une approximation de Taylor de premier ordre sur la fonction $h(\theta)$. Alors, on a que

$$h(\hat{\theta}) \approx h(\theta) + \frac{\partial}{\partial \theta} h(\theta) (\hat{\theta} - \theta)$$

On a que

$$E[h(\hat{\theta})] = h(\hat{\theta})$$

et

$$\begin{aligned}\text{Var}(\hat{\theta}) &= \text{Var}\left(h(\theta) + \frac{\partial}{\partial \theta} h(\theta)(\hat{\theta} - \theta)\right) \\ &= \left(\frac{\partial}{\partial \theta} h(\theta)\right)^2 \text{Var}(\hat{\theta})\end{aligned}$$

Dans un contexte multivarié (où il faut estimer plus d'un paramètre), on a

$$\text{Var}(h(\hat{\theta})) = \mathbf{h}^\top I(\theta)^{-1} \mathbf{h}$$

où \mathbf{h} est le vecteur des dérivées partielles de $h(\theta)$:

$$\mathbf{h} = \begin{bmatrix} \frac{\partial}{\partial \theta_1} h(\theta) \\ \frac{\partial}{\partial \theta_2} h(\theta) \\ \dots \\ \frac{\partial}{\partial \theta_k} h(\theta) \end{bmatrix}$$

8.5 Likelihood Ratio Test (LRT)

Le test du rapport de vraisemblance (*Likelihood ratio test*) permet de valider si la fonction de vraisemblance d'un modèle avec paramètres *contraints* est proche de la fonction de vraisemblance maximisée lorsqu'il n'y a aucune contrainte sur les paramètres. Ce test a plusieurs applications, notamment ici et dans l'ajustement de modèles linéaires généralisés. Une application directe dans la ?? est présentée. En résumé, on la mesure statistique du test est la suivante :

$$\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$$

où

- › $L(\theta_0)$ représente la fonction de vraisemblance du modèle avec paramètres contraints ;
- › $\ell(\theta)$ représente la fonction de vraisemblance pour le modèle complet.

On peut prouver que

$$-2 \ln \lambda \sim \chi_{dl_1 - dl_0}^2 \quad (8.11)$$

où

- › dl_1 représente le nombre de degrés de libertés du modèle complet (i.e. le nombre de paramètres qu'on peut modifier, qui ne sont pas fixés)
- › dl_0 le nombre de paramètres non fixés du modèle avec paramètres contraints à tester

où et .

Dans le contexte d'estimation de paramètres, on aura le modèle complet qui correspond en fait

Souvent, l'?? est plutôt présentée comme

$$2 (\ell(\theta) - \ell(\theta_0)) \sim \chi_{dl_1 - dl_0}^2 \quad (8.12)$$

Le test LRT le ratio λ sera toujours ≤ 1 , car le modèle avec paramètres estimés ne peut pas estimer parfaitement la distribution. Toutefois, on veut que la différence entre la logvraisemblance de la distribution ($\ell(\theta)$) et celle du modèle avec paramètres estimés ($\ell(\theta_0)$) ne soit pas significativement trop élevée⁶, i.e.

$$2 (\ell(\theta) - \ell(\theta_0)) \leq \chi_{dl_1 - dl_0, 1 - \alpha}^2 \quad (8.13)$$

Construction d'un intervalle de confiance Comme n'importe quel test d'hypothèse, il est possible avec le LRT de construire un intervalle de confiance autour de nos paramètres estimés en *inversant* le test. En effet, $\hat{\theta}_0$ doit être le paramètre qui respecte l'?? à un seuil $1 - \alpha$.

Exemple 8.5.1 Exemple avec Q13.57 du Loss Models



13.57 The following 20 observations were collected. It is desired to estimate $\Pr(X > 200)$. When a parametric model is called for, use the single-parameter Pareto distribution for which $F(x) = 1 - (100/x)^\alpha$, $x > 100$, $\alpha > 0$.

132	149	476	147	135	110	176	107	147	165
135	117	110	111	226	108	102	108	227	102

6. Autrement dit, on va rejeter H_0 que le modèle avec les paramètres estimés est adéquat si la statistique λ dépasse le quantile χ^2 .

On sait que $\hat{\alpha} = 2,848$ et $\hat{\theta} = 100$.

Trouver l'intervalle de confiance à 95% en inversant le test du ratio de vraisemblance (LRT).

(1) On peut trouver que la fonction de densité est

$$\begin{aligned} f(x; \alpha) &= \frac{\partial}{\partial \alpha} F(x; \alpha) \\ &= \frac{\partial}{\partial \alpha} \left(1 - \left(\frac{100}{x} \right)^\alpha \right) \\ &= \frac{\alpha 100^\alpha}{x^{\alpha+1}} \end{aligned}$$

(2) Les fonctions de vraisemblance L et log-vraisemblance ℓ sont, respectivement

$$\begin{aligned} L(\alpha) &= \prod_{i=1}^{20} f(x_i; \alpha) \\ &= \prod_{i=1}^{20} \frac{\alpha 100^\alpha}{x_i^{\alpha+1}} \end{aligned}$$

et

$$\begin{aligned} \ell(\alpha) &= \ln L(\alpha) \\ &= \sum_{i=1}^{20} \ln(\alpha) + \alpha \ln(100) - (\alpha + 1) \ln(x_i) \\ &= 20 \ln(\alpha) + 20\alpha \ln(100) - (\alpha + 1) \sum_{i=1}^{20} \ln(x_i) \end{aligned}$$

Avec la fonction DATA de la calculatrice TI-30XS, on peut facilement calculer le terme $\sum_{i=1}^{20} \ln x_i = 99,125$. En remplaçant α par $\hat{\alpha} = 2,848$, on obtient donc

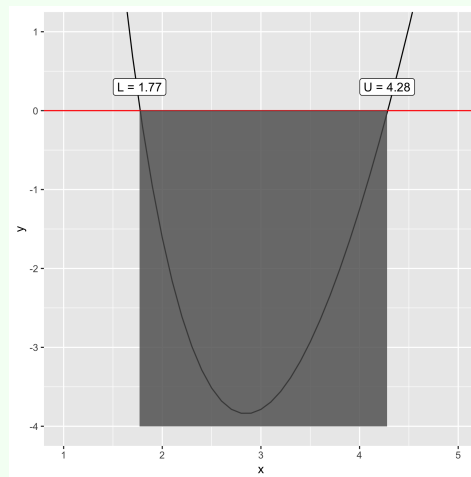
$$\ell(\alpha) = -98,190166$$

(3) On peut trouver $\hat{\alpha}_0$ qui résoud

$$\begin{aligned}
 2(\ell(\hat{\alpha}) - \ell(\hat{\alpha}_0)) &\leq \chi_{1,0.95}^2 \\
 2(-98,190166 - (20 \ln(\hat{\alpha}_0) + \\
 20\hat{\alpha}_0 \ln(100) - (\hat{\alpha}_0 + 1)99,125) &\leq 3,841 \\
 -196,38033 - 40 \ln \hat{\alpha}_0 - 40\hat{\alpha}_0 \ln 100 + 198,25(\hat{\alpha}_0 + 1) &\leq 3,841 \\
 -200,2213 - 40 \ln \hat{\alpha}_0 - 40\hat{\alpha}_0 \ln 100 + 198,25(\hat{\alpha}_0 + 1) &\leq 0
 \end{aligned}$$

Avec la fonction TABLE de la calculatrice TI-30XS, on peut entrer cette fonction, et tester pour des valeurs de $\hat{\alpha}_0$ jusqu'à résolution. On doit obtenir une borne $L < \hat{\alpha}$ et une borne $U > \hat{\alpha}$.

(4) Graphiquement,



(5) Alors,

$$\alpha \in [1,77; 4,28]$$

8.6 Exercices recommandés

Les exercices recommandés dans le *Loss Models, 4th edition* pour ce chapitre sont les suivants :

- › Méthode des moments et percentiles : 13.1 à 13.22
- › *MLE* : 13.30, 13.32, 13.33, 13.37, 13.38 à 13.40, 13.45 à 13.51 et 13.54 à 13.57
- › Variance et intervalle de confiance : 13.62 à 13.72
- › Non-normal confidence interval : 13.74

Chapitre 9

Sélection de modèles (16)

9.1 Introduction

Il existe plusieurs tests (graphiques, tests d'hypothèses, etc.) pour valider si un modèle est adéquat. Les 2 tests vus dans le cadre du cours sont le *Goodness-of-fit* et le *Likelihood ratio test*. Souvent, certaines valeurs de ces tests peuvent aussi être utilisées pour comparer différents modèles entre-eux.

9.2 *Goodness-of-fit* Chi-Square test

Avec le test *Goodness-of-fit*, on cherche à valider que les *prédictions* du modèle ne sont pas trop différentes des valeurs empiriques observées. On valide le tout avec une mesure très similaire à l'erreur quadratique, dénotée X^2 :

$$X^2 = \frac{\sum_{i=1}^k n(\hat{p}_i - p_{ni})^2}{\hat{p}_i}$$

où

- › \hat{p}_i est la probabilité (selon le modèle) d'observer une valeur dans la i^{e} classe ;
- › p_{ni} est la probabilité empirique observée d'observations dans la i^{e} classe ;
- › n est le nombre d'observations.

Toutefois, le *Loss Models* utilise toujours la formule suivante pour les exercices :

$$X^2 = \frac{\sum_{j=1}^k (E_j - O_j)^2}{E_j} \quad (9.1)$$

où $E_j = n\hat{p}_i$ est le nombre de valeurs qu'on s'attend à avoir dans la i^e classe et $O_j = np_{ni}$ le nombre d'observations dans la i^e classe.

De plus, on peut prouver que

$$X^2 \sim \chi_{k-p-1}^2$$

où k est le nombre de classes (on généralise pour le cas où les données seraient groupés) et p le nombre de paramètres qui varient dans le modèle. Formellement, le test d'hypothèse est :

$$H_0 : X^2 \leq \chi_{k-p-1}^2$$

$$H_1 : X^2 > \chi_{k-p-1}^2$$

9.3 Likelihood Ratio Test (LRT)

En utilisant la statistique du Chi-carré, on cherche à valider que la différence entre la fonction de log-vraisemblance du modèle complet versus le modèle proposé n'est pas significativement trop élevée, sans quoi on devra rejeter H_0 , et alors le modèle avec contraintes ne sera pas une bonne simplification du modèle complet. On teste la *significativité* avec la statistique T telle que

$$T = 2(L_1 - L_0) \sim \chi_{dl=k_1-k_0}^2 \quad (9.2)$$

où $L_1 = \ell(\theta_{(H_1)})$ et $L_0 = \ell(\theta_{(H_0)})$, k_1 est le nombre de paramètres *libres* (i.e. qu'on peut modifier) du modèle complet et k_0 le nombre de paramètres libres du modèle avec contraintes.

9.4 Critères de sélection d'un modèle

L'expérience compte pour beaucoup dans la sélection du bon modèle. Il y a toutefois quelques critères et statistiques qu'on peut analyser et comparer entre les modèles pour faire certains choix : on va préférer un modèle avec

1. la plus **faible valeur** pour le test **Kolmogorov-Smirnov** ;
2. la plus **faible valeur** pour le test **Anderson-Darling** ;
3. la plus **faible valeur** pour le test **Goodness-of-fit** ;
4. la plus **haute valeur** pour la *p-value* du test **Goodness-of-fit** ;
5. la plus **haute valeur** pour la **fonction de vraisemblance à son maximum**.

9.5 Exercices recommandés

Test Kolmogorov-Smirnov 16.4 et 16.5¹

Goodness-of-fit test et LRT 16.10 à 16.15

Critères de sélection 16.23 et 16.24

1. Cette matière n'est pas testée à l'examen final A2018 en IARD1.

Chapitre 10

Estimation Bayésienne (15)

10.1 Définition de l'estimateur Bayésien

Définition 10.1.1 Distribution *a priori*



Soit un paramètre θ d'une distribution quelconque. Afin de réaliser une estimation Bayésienne, on connaît *a priori* la distribution que prend le paramètre θ , qu'on dénote par $\pi(\theta)$.

Alors, notre distribution des pertes est conditionné par rapport à la valeur que θ prend (i.e. $f_{X|\Theta}$).

Définition 10.1.2 Distribution *a posteriori*



La distribution *a posteriori* nous permet de savoir avec quelle probabilité non-nulle notre paramètre θ peut prendre une certaine valeur, sachant qu'on a observé certains x , qu'on dénote comme $\pi_{\Theta|X}(\theta|x)$:

$$\pi_{\Theta|X}(\theta|x) = \frac{f_{\Theta,X}(\theta,x)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta)\pi(\theta)}{\int f_{X|\Theta}(x|\theta)\pi(\theta)d\theta} \quad (10.1)$$

L'idée est de remplacer les différentes distributions dans l'??, et en déduire une distribution avec une paramétrisation différente^a.

^a. Souvent, la distribution *a posteriori* aura la même distribution que celle *a priori*, mais avec des paramètres différents.

L'estimateur Bayésien L'estimateur Bayésien est défini comme l'espérance du paramètre θ , sachant la distribution de X . En d'autres mots, on veut l'espérance de la distribution *a posteriori* :

$$\hat{\theta}_{BAYES} = E [\Theta | X] \quad (10.2)$$

Propriétés de l'estimateur Voici quelques propriétés importantes de l'estimateur :

- › $\hat{\theta}_{BAYES}$ est biaisé (comparativement à l'estimateur MLE qui est non-biaisé);
- › Par contre, l'estimateur Bayésien a une plus petite variance que l'estimateur MLE. On sacrifie donc le Biais pour diminuer la variance de l'estimateur;
- › **Remarque** : Si n est grand, il peut être préférable d'utiliser l'estimateur MLE.

10.1.1 Estimation par simulation

Une fois qu'on a développé et trouver la distribution *a posteriori* de $\Theta | X$ (??), alors on peut facilement simuler l'estimateur Bayésien :

1.

Exemple 10.1.1 Exemple avec la Loi de Poisson



Soit $X | \Lambda \sim \text{Pois}(\Lambda)$ et $\Lambda \sim \Gamma(\alpha, \theta)$. Alors, on a la distribution *a priori* suivante :

$$\Lambda \sim \pi(\lambda) = \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha} e^{-\lambda/\theta}$$

Évidemment, on va s'attendre à ce que la distribution *a posteriori* obéisse à une loi Gamma aussi. On a (avec $c(x)$ une constante) :

$$\begin{aligned} \pi_{\Lambda|X}(\lambda|x) &= c(x) \frac{\lambda^{\alpha-1} e^{-\lambda/\theta}}{\Gamma(\alpha)\theta^\alpha} \frac{\lambda^\alpha e^{-\lambda}}{x!} \\ &= k(x) \lambda^{x+\alpha-1} e^{-\lambda(\frac{1}{\theta} + 1)} \end{aligned}$$

avec $c(x)$ et $k(x)$ des constantes uniquement fonction de x .

On trouve donc que $\pi_{\Lambda|X}(\lambda|x) \sim \Gamma\left(\alpha' = x + \alpha, \beta' = \left(\frac{1}{\theta} + 1\right)^{-1}\right)$
 Puisque l'estimateur Bayésien est l'espérance de la distribution *a posteriori*, on a

$$\begin{aligned}\hat{\lambda}_{BAYES} &= E[\Lambda|X] \\ &= \int \lambda \cdot \pi_{\Lambda|X}(\lambda|x) d\lambda \\ &= \alpha' \beta' \\ &= (x + \alpha) \left(\frac{\theta}{1 + \theta} \right)\end{aligned}$$

Exemple 10.1.2 Cas plus général ...



On a le modèle suivant pour X , tel que

$$f_{X_1, \dots, X_n|\Lambda}(x_1, \dots, x_n|\lambda) = \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \dots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$$

Ainsi que la distribution *a priori* telle que

$$\Lambda \sim \pi(\lambda) = \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha} e^{-\lambda/\theta}$$

Alors, la distribution *a posteriori* est donc

$$\begin{aligned}f_{\Lambda|X}(\lambda|x) &= c(x) \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha} e^{-\lambda/\theta} \cdot \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \dots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!} \\ &= k(x) \lambda^{\sum x_i + \alpha - 1} e^{-\lambda/\theta} e^{-n\lambda} \\ &= k(x) \lambda^{\sum x_i + \alpha - 1} e^{-\lambda(\frac{1}{\theta} + n)}\end{aligned}$$

On déduit donc que la distribution *a posteriori* obéit à une loi Gamma de paramètres $\alpha' = \sum x_i + \alpha$ et $\beta' = \left(\frac{1}{\theta} + n\right)^{-1} = \frac{\theta}{1+n\theta}$. Alors, on

déduit facilement l'estimateur Bayésien :

$$\begin{aligned}
 \hat{\lambda}_{BAYES} &= \alpha' \beta' \\
 &= \frac{\theta (\sum_{i=1}^n x_i + \alpha)}{1 + n\theta} \\
 &= \frac{n\theta}{1 + n\theta} \left(\frac{\sum_{i=1}^n x_i}{n} \right) + \frac{1}{1 + n\theta} (\alpha\theta) \\
 &= \frac{n\theta}{1 + n\theta} \hat{\lambda}_{MLE} + \frac{1}{1 + n\theta} \underbrace{(\alpha\theta)}_{\text{Moy. distr. a priori}}
 \end{aligned}$$

Remarque On peut donc voir l'estimateur Bayésien comme une moyenne pondérée entre l'estimateur du maximum de vraisemblance et la moyenne de la distribution *a priori*. Lorsque $n \rightarrow \infty$, l'estimateur Bayésien sera semblable à l'EMV.

Exemple 10.1.3 Exemple avec la loi Normale



On prend le modèle $Y|\Theta = \frac{\sum_{i=1}^n x_i}{n} \sim N\left(\theta, \frac{\sigma^2}{n}\right)$, avec σ^2 connu. La densité *a priori* obéit aussi à une loi normale, telle que $\Theta \sim N(\theta_0, \sigma_0^2)$, où θ_0, σ_0^2 sont connus. On va donc s'attendre à avoir une distribution *a posteriori* qui va suivre une loi normale aussi. Alors, la distribution *a posteriori* $\pi_{\Theta|X}(\theta|x)$ peut être trouvée :

$$\pi_{\Theta|X}(\theta|x) = c(y) e^{\frac{-(y-\theta)^2}{2(\sigma^2/n)}} e^{\frac{-(\theta-\theta_0)^2}{2(\sigma_0^2)}}$$

On va essayer de jumeler les 2 exposants des e , pour retrouver une certaine forme de la loi normale :

$$\frac{(y - \theta)^2}{2(\sigma^2/n)} + \frac{(\theta - \theta_0)^2}{2(\sigma_0^2)} = \frac{\sigma_0^2(y - \theta)^2 + \frac{\sigma^2}{n}(\theta - \theta_0)^2}{2\frac{\sigma^2}{n}\sigma_0^2}$$

Si on développe les termes séparément :

$$\sigma_0^2(y - \theta)^2 = \sigma_0^2(y^2 + \theta^2 - 2\theta y)$$

et

$$\frac{\sigma^2}{n}(\theta - \theta_0)^2 = \frac{\sigma^2}{n}(\theta^2 + \theta_0^2 - 2 \cdot \theta \cdot \theta_0)$$

Alors,

$$\sigma_0^2(y - \theta)^2 + \frac{\sigma^2}{n}(\theta - \theta_0)^2 = \left(\sigma_0^2 + \frac{\sigma^2}{n}\right) \theta^2 - 2 \left(y\sigma_0^2 + \theta_0 \frac{\sigma^2}{n}\right) \theta$$

Alors, la distribution *a posteriori* peut être ré-écrite comme

$$\pi_{\Theta|X}(\theta|x) = k(y) \exp \left\{ -\frac{1}{2} \left(\theta^2 - 2 \left(\frac{y\sigma_0^2 + \theta_0 \frac{\sigma^2}{n}}{\sigma_0^2 + \frac{\sigma^2}{n}} \right) \theta / \frac{\frac{\sigma^2}{n} \sigma_0^2}{\sigma_0^2 + \frac{\sigma^2}{n}} \right) \right\}$$

Si on manipule le numérateur, on trouve

$$\pi_{\Theta|X}(\theta|x) = k(y) \exp \left\{ -\frac{1}{2\gamma} \left(\theta - \frac{y\sigma_0^2 + \theta_0 \frac{\sigma^2}{n}}{\sigma_0^2 + \frac{\sigma^2}{n}} \right)^2 \right\}$$

Donc

$$\pi_{\Theta|X}(\theta|x) \sim N(\mu', \sigma' = \gamma)$$

où $\mu' = \frac{y\sigma_0^2 + \theta_0 \frac{\sigma^2}{n}}{\sigma_0^2 + \frac{\sigma^2}{n}}$. On peut aussi le ré-écrire (aussi appelé estimateur de crédibilité, sera vu en IARD2) :

$$\hat{\theta}_{BAYES} = \left(\frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma^2}{n}} \right) \bar{x} + \left(\frac{\sigma^2/n}{\sigma_0^2 + \frac{\sigma^2}{n}} \right) \theta_0$$

10.2 Exercices recommandés

Les exercices recommandés dans le *Loss Models, 4th edition* pour ce chapitre sont les suivants :

‣ 15.14 a 15.18, 15.21 a 15.27

.

Troisième partie

Annexes

Annexe A

Jeux de données du Loss Models

Voici les jeux de données qui sont utilisés dans le Loss Models, qui sont constamment utilisés dans les exercices du livre.

A.1 Data set A

Table 11.1 Data Set A.	
Number of accidents	Number of drivers
0	81,714
1	11,306
2	1,618
3	250
4	40
5 or more	7

A.2 Data set B

Table 11.2 Data Set B.

27	82	115	126	155	161	243	294	340	384
457	680	855	877	974	1,193	1,340	1,884	2,558	15,743

A.3 Data set C

Table 11.3 Data Set C.

Payment range	Number of payments
0–7,500	99
7,500–17,500	42
17,500–32,500	29
32,500–67,500	28
67,500–125,000	17
125,000–300,000	9
Over 300,000	3

A.4 Data set D1

Table 11.4 Data Set D1.

Policyholder	Time of death	Time of surrender
1	–	0.1
2	4.8	0.5
3	–	0.8
4	0.8	3.9
5	3.1	1.8
6	–	1.8
7	–	2.1
8	–	2.5
9	–	2.8
10	2.9	4.6
11	2.9	4.6
12	–	3.9
13	4.0	–
14	–	4.0
15	–	4.1
16	4.8	–
17	–	4.8
18	–	4.8
19–30	–	–

A.5 Data set D2

Table 11.5 Data Set D2.

Policy	First observed	Last observed	Event ^a	Policy	First observed	Last observed	Event
1	0	0.1	s	16	0	4.8	d
2	0	0.5	s	17	0	4.8	s
3	0	0.8	s	18	0	4.8	s
4	0	0.8	d	19-30	0	5.0	e
5	0	1.8	s	31	0.3	5.0	e
6	0	1.8	s	32	0.7	5.0	e
7	0	2.1	s	33	1.0	4.1	d
8	0	2.5	s	34	1.8	3.1	d
9	0	2.8	s	35	2.1	3.9	s
10	0	2.9	d	36	2.9	5.0	e
11	0	2.9	d	37	2.9	4.8	s
12	0	3.9	s	38	3.2	4.0	d
13	0	4.0	d	39	3.4	5.0	e
14	0	4.0	s	40	3.9	5.0	e
15	0	4.1	s				

^a“s” indicates surrender, “d” indicates death, “e” indicates expiration of the 5-year period.

Annexe B

Rappels mathématiques

B.1 Algèbre linéaire

Soit la matrice A , telle que

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Matrice transposée la matrice transposée est définie par A^\top , telle que

$$A^\top = \begin{bmatrix} a & -c \\ -b & d \end{bmatrix}$$

Déterminant d'une matrice On peut calculer le déterminant $\det(A)$ de la matrice A tel que

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Inverse d'une matrice L'équivalent de l'opération $\frac{1}{A}$ en algèbre linéaire est de calculer la matrice inverse de A^{-1} , telle que

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} a & -c \\ -b & d \end{bmatrix}$$

où on multiplie par la matrice adjointe de A . Il faut normalement calculer les cofacteurs, mais le cas à 2 dimensions est un cas simplifié.