

Mathématiques actuarielles IARD-1
ACT-2005
Notes de cours

Gabriel Crépeault-Cauchon
Nicholas Langevin

14 novembre 2018

Table des matières

3	Quantités des distributions à connaître	1
3.1	Moments	1
3.4	Queue de distribution	4
3.4.1	Classification selon les moments	4
3.4.2	Classification selon les comportements limites des ailes de distribution	4
3.4.3	Classification basée sur la fonction de Hazard	5
8	Fréquence et sévérité avec modifications aux contrats	6
8.2	Deductibles	6
8.2.1	Ordinary deductible	6
8.2.2	Franchise deductible	8
8.3	Loss Elimination Ratio	9
8.4	Policy Limits	11
8.4.1	Définitions	11
8.4.2	Fonctions reliées	11
8.5	Coassurance, deductible et limites	12
11	Estimation de données complètes	13
11.2	Estimation de données complètes	13
11.2.1	Estimation de la fonction de répartition empirique	13
11.2.2	<i>Cumulative hazard-rate function</i>	14
11.2.3	Notation à utiliser pour la distribution empirique	15
11.2.4	Estimateur de Nelson Åalen	15
11.3	Distribution empirique avec données groupées	16
11.3.1	Fonction OGIVE	16

12	Estimation de données modifiées	18
12.1	Point estimation	18
12.1.1	Définitions importante	18
12.2	Espérance, Variance et et intervalle d'estimation	20

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.

Chapitre 3

Quantités des distributions à connaître

3.1 Moments

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

$$\mu'_k = E[X^k] \quad (3.1)$$

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

$$\mu_k = E[(X - \mu)^k] \quad (3.2)$$

Exemple 3.1.1 Quelques exemples de moments centraux



La variance est le 2^e moment central :

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

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

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

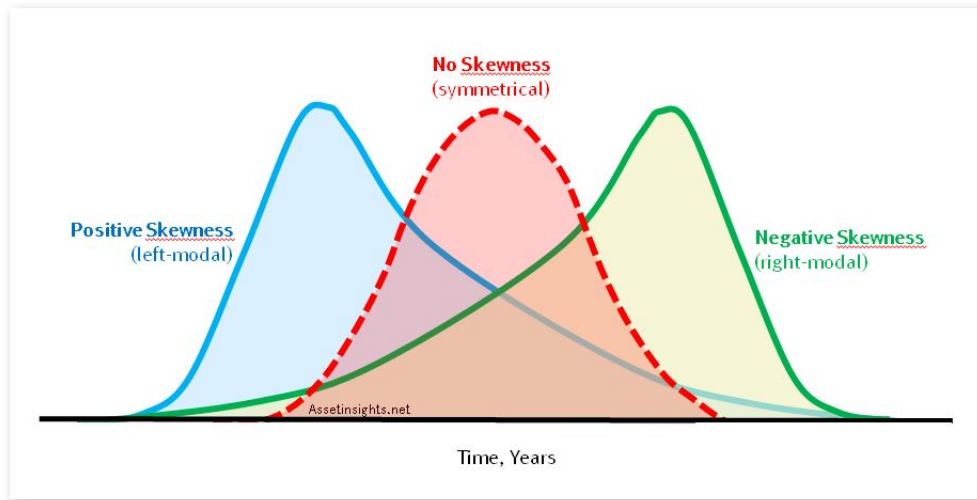
Coefficient de variation

$$CV = \frac{\sigma}{E[X]} \quad (3.3)$$

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

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \quad (3.4)$$

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

$$\text{Kurtosis} = \frac{\mu_4}{\sigma^4} \quad (3.5)$$

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 :

$$e(d) = E[Y^P] = E[X - d | X \geq d] = \frac{\int_d^\infty S_X(x)}{S_X(d)} \quad (3.6)$$

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 :

$$E[Y^L] = E[(X - d)_+] = \int_d^{\infty} (x - d)f_X(x)dx \quad (3.7)$$

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

$$E[X \wedge u] = \int_0^u f_X(x)dx \quad (3.8)$$

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

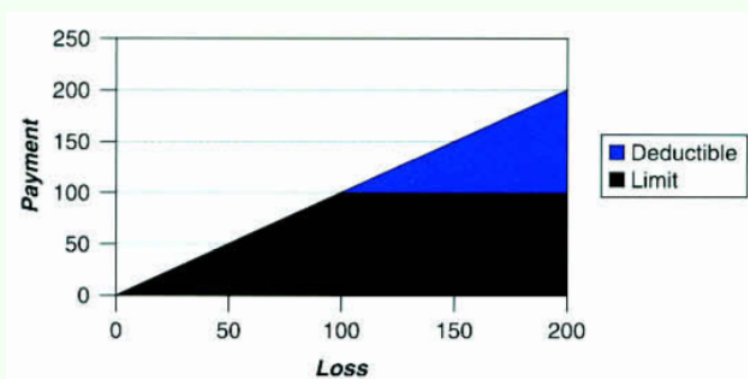
$$E[X \wedge u] = \int_0^u S_X(x)dx$$

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

Exemple 3.1.2 lien entre le déductible et la limite



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



3.4 Queue de distribution

3.4.1 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*)

3.4.2 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 3.4.1

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

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

La pareto a une queue plus lourde que la gamma

3.4.3 Classification basée sur la fonction de Hazard

Définition 3.4.1 Hazard rate function



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

$$h_X(x) = \frac{f_X(x)}{S_X(x)} \quad (3.9)$$

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.

Chapitre 8

Fréquence et sévérité avec modifications aux contrats

8.2 Déductibles

8.2.1 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*,

$$Y^L = (X - d)_+ = \begin{cases} 0 & , X \leq d \\ X - d & , X > d \end{cases} \quad (8.1)$$

$$Y^P = (X - d)_+ = \begin{cases} \text{Non-défini} & , X \leq d \\ X - d & , X > d \end{cases} \quad (8.2)$$

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

$$E[Y^L] = E[(X-d)_+] = E[X] - E[X \wedge d] \quad (8.3)$$

$$E[Y^P] = \frac{E[(X-d)_+]}{S_X(d)} = \frac{E[X] - E[X \wedge d]}{S_X(d)} \quad (8.4)$$

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

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

8.2.2 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 ,

$$Y^P = \begin{cases} \text{non-défini} & X \leq d \\ X & , X > d \end{cases} \quad (8.5)$$

Fonctions reliées

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

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

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.

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

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

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

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

$$E[Y^L] = E[(X - d)_+^F] = E[X] - E[X \wedge d] + dS_X(d) \quad (8.6)$$

$$E[Y^P] = E[(X - d)_+^F] = \frac{E[X] - E[X \wedge d]}{S_X(d)} + d \quad (8.7)$$

8.3 Loss Elimination Ratio

Définition 8.3.1 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,

$$LER = \frac{E[X \wedge d]}{E[X]} \quad (8.8)$$

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,

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

□

8.4 Policy Limits

8.4.1 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

$$Y = \begin{cases} X & , X \leq u \\ u & , Y > u \end{cases} \quad (8.9)$$

8.4.2 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.

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

□

8.5 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 ,

$$Y^L = \begin{cases} 0 & , x < \frac{d}{1+r} \\ \alpha \left((1+r)x - d \right) & , \frac{d}{1+r} \leq x < \frac{u}{1+r} \\ \alpha(u - d) & , x \geq \frac{u}{1+r} \end{cases} \quad (8.10)$$

et pour Y^P ,

$$Y^L = \begin{cases} \text{Non-défini} & , x < \frac{d}{1+r} \\ \alpha \left((1+r)x - d \right) & , \frac{d}{1+r} \leq x < \frac{u}{1+r} \\ \alpha(u - d) & , x \geq \frac{u}{1+r} \end{cases} \quad (8.11)$$

Espérance

$$E[Y^L] = \alpha(1+r) \left(E \left[X \wedge \frac{u}{1+r} \right] - E \left[X \wedge \frac{d}{1+r} \right] \right) \quad (8.12)$$

$$E[Y^P] = \frac{E[Y^L]}{S_X\left(\frac{d}{1+r}\right)} \quad (8.13)$$

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

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.

Chapitre 11

Estimation de données complètes

11.2 Estimation de données complètes

11.2.1 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 $F(t)$:

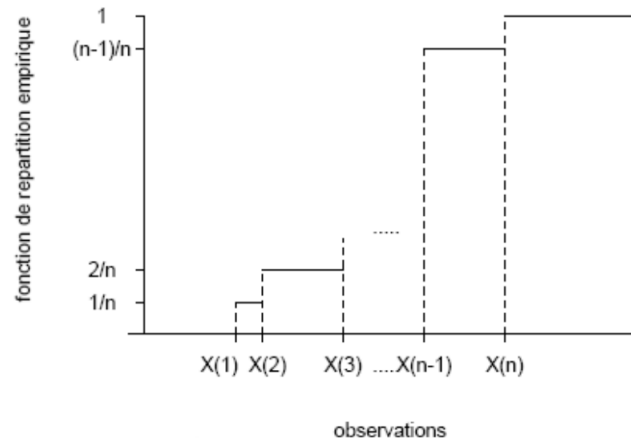
$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i \leq t\}} \quad (11.1)$$

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

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

$$F_n(t) = \begin{cases} 0 & , t < x_{(1)} \\ \frac{1}{n} & , x_{(1)} \leq t < x_{(2)} \\ \frac{2}{n} & , x_{(2)} \leq t < x_{(3)} \\ \dots & \\ 1 & , t \geq x_{(n)} \end{cases} \quad (11.2)$$

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,

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

11.2.2 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

$$H_X(x) = -\ln S_X(x) \quad (11.3)$$

11.2.3 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 :

$$F_n(x) = \begin{cases} 0 & , x < y_1 \\ 1 - \frac{r_j}{n} & , y_{j-1} \leq x \leq y_j \quad j = 2, \dots, k \\ 1 & , x \geq y_k \end{cases} \quad (11.4)$$

11.2.4 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* :

$$\hat{H}(x) = \begin{cases} 0 & , x < y_1 \\ \sum_{i=1}^{j-1} \frac{s_i}{r_i} & , y_{j-1} \leq x \leq y_j \quad j = 2, \dots, k \\ \sum_{i=1}^k \frac{s_i}{r_i} & , x \geq y_k \end{cases} \quad (11.5)$$

11.3 Distribution empirique avec données groupées

11.3.1 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 OGIV ci-dessous :

$$F_n^{\text{OGIVE}}(x) = \frac{c_j - x}{c_j - c_{j-1}} F_n(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}} F_n(c_j) \quad (11.6)$$

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 (11.6), on obtient

$$\begin{aligned} f_n(x) &= \frac{\partial}{\partial x} F_n(x) \\ &= \frac{1}{c_j - c_{j-1}} F_n(c_j) - \frac{1}{c_j - c_{j-1}} F_n(c_{j-1}) \\ &= \frac{F_n(c_j) - F_n(c_{j-1})}{c_j - c_{j-1}} \end{aligned} \tag{11.7}$$

Exemple 11.3.1 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)

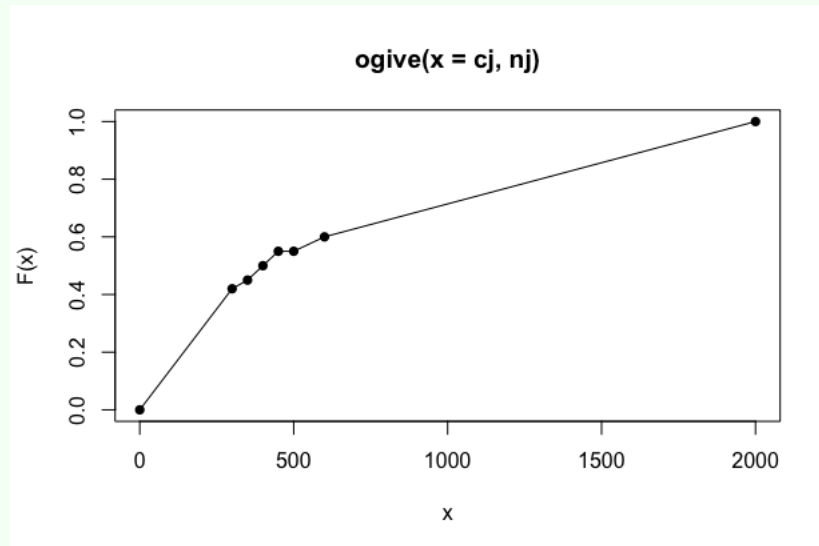
## On définit le vecteur cj des bornes et
## un vecteur nj des fréquences empiriques
# note : il faut forcer une borne supérieure
cj <- c(0, 300, 350, 400, 450, 500, 600, 2000)
nj <- c(42, 3, 5, 5, 0, 5, 40)

# 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 :



Chapitre 12

Estimation de données modifiées

12.1 Point estimation

12.1.1 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 11.2.4. Avec l'information des données et en utilisant la notation vue à la sous-section précédente, on peut estimer la fonction de survie empirique :

$$\hat{S}_n(t) = \begin{cases} 1 & , 0 \leq t < y_1 \\ \prod_{i=1}^{j-1} \left(\frac{r_i - s_i}{r_i} \right) & , y_{j-1} \leq t < y_j \quad j = 2, \dots, k \\ \prod_{i=1}^k \left(\frac{r_i - s_i}{r_i} \right) & t \geq y_k \end{cases} \quad (12.1)$$

12.2 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