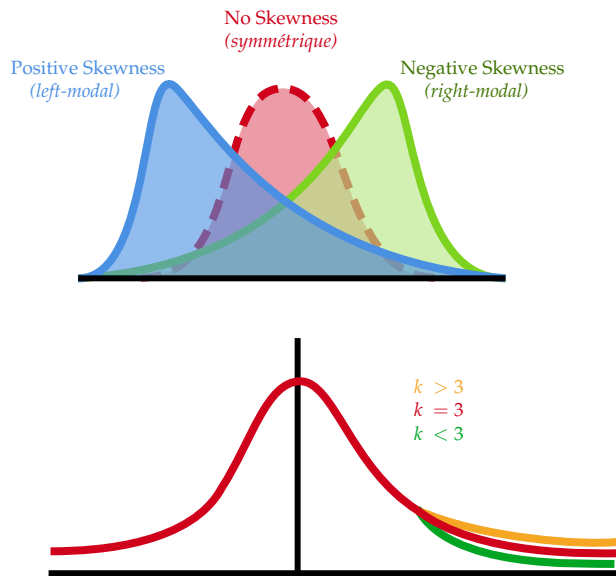


3 Estimation non-paramétrique

Moments à savoir

$$\begin{aligned}\mu'_k &= E[X^k] & \mu_k &= E[(X - \mu)^k] \\ CV &= \frac{\sigma}{\mu} \\ \text{Skewness : } \gamma &= \frac{\mu'_3}{\sigma^3} & \text{Kurtosis : } \kappa &= \frac{\mu'_4}{\sigma^4}\end{aligned}$$



3 critères pour évaluer les queues de distributions

1. La loi avec le moins de moments a la queue la plus lourde.
2. Première à diverger du quotient des distributions a la queue la plus lourde.

$$\lim_{x \rightarrow \infty} \frac{f_{X_1}(x)}{f_{X_2}(x)}$$

3. Si la fonction hasard $h(x) = \frac{f_X(x)}{S_X(x)}$ est croissante alors

la queue est fine, sinon elle est lourde.

$$\begin{aligned}h'_X(x) &< 0 & \text{queue lourde} \\ h'_X(x) &> 0 & \text{queue fine}\end{aligned}$$

Quantités des distributions à connaître

Y^P : **Excess loss**, alias **left truncated** and **shifted** variable.
On interprète comme le *montant de perte en excès d'un déductible d* sachant que la perte est au delà de ce montant.

Y^L : **Left censored** and **shifted** variable.
Elle est défini comme étant 0 pour toutes les pertes inférieures à d , alors que l'excès-moyen n'est simplement pas défini dans ces cas.
Donc, celle-ci a une masse à 0.

Y : **Limited loss**, alias **right censored** variable.

shifted : d est soustrait des valeurs restantes.

On peut visualiser le déplacement de la courbe de densité à la gauche.

left truncated : Toutes valeurs inférieures à d ne sont pas observées.

left censored : Toutes valeurs inférieures à d sont égales à 0.

right censored : Toutes valeurs supérieures à u sont égales à u .

Pour exemple, lorsqu'il y a une limite sur une police d'assurance les valeurs au-delà ne sont pas typiquement inscrites à leur vrai montant, mais plutôt comme la limite u .

Moments

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

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

$$E[Y] = E[X \wedge d] = \int_0^d f_X(x) dx \Leftrightarrow \int_0^d S_X(x) dx$$

8 Fréquence et sévérité avec modifications aux contrats

Déductible ordinaire

L'assureur paye tout montant en excédent du montant d .

$$Y_{(O)}^{(L|P)} = (X - d)_+ = \begin{cases} (0 | \text{non-défini}) & , X \leq d \\ X - d & , X > d \end{cases}$$

Déductible franchise

L'assureur paye l'entièreté des coûts pour toute perte qui surpasse le montant d .
Pour éviter les petites réclamations

$$Y_{(F)}^{(L|P)} = (X - d)_+ = \begin{cases} (0 | \text{non-défini}) & , X \leq d \\ X & , X > d \end{cases}$$

Moments

$$E[Y_{(O|F)}^{(L|P)}] = \frac{E[X] - E[X \wedge d] + dS_X(d)}{S_X(d)}$$

De plus, on note que :

$$E[Y_{(O)}^{(P)}] = e_X(d)$$

$$E[Y_{(O)}^{(L)}] = \pi_X(d)$$

Fonctions

$$f_{Y_{(O)}^{(L|P)}} = \frac{f_X(y + d)}{S_X(d)}$$

$$h_{Y_{(O)}^{(L|P)}} = h_X(y + d)$$

$$S_{Y_{(O)}^{(L|P)}} = \frac{S_X(y + d)}{S_X(d)}$$

$$F_{Y_{(O)}^{(L|P)}} = \frac{F_X(y + d) - F_X(d)}{S_X(d)}$$

LER et inflation du déductible ordinaire

Le LER nous donne le pourcentage de perte qu'on ne paie pas grâce au déductible

$$\begin{aligned} LER &= \frac{E[X] - E[(X - u)_+]}{E[X]} \\ &= \frac{E[X \wedge u]}{E[X]} \end{aligned}$$

Soit $X^I = (1 + r)X$

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

$$f_{X^I}(x) = \frac{f_X\left(\frac{y}{1+r}\right)}{1+r}$$

$$F_{X^I}(x) = F_X\left(\frac{y}{1+r}\right)$$

Limite de police

L'assureur paye un maximum de u

$$Y = (X \wedge u) = \begin{cases} X, & X < u \\ u, & X \geq u \end{cases}$$

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

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

Coassurance

L'assureur paye une fraction, α , de la perte.

Si la coassurance est la seule modification, alors nous obtenons $Y = \alpha X$.

L'impact sur les fonctions est le même qu'avec de l'inflation.

Formule récapitulative

Lorsque les 4 items sont présent (déductible ordinaire, limite, inflation et coassurance.

$$\begin{aligned} Y_{(O)}^{(L|P)} &= \begin{cases} (0|\text{Non-défini}) & , x < \frac{d}{1+r} \\ \alpha((1+r)x - d) & , \frac{d}{1+r} \leq x < \frac{u}{1+r} \\ \alpha(u - d) & , x \geq \frac{u}{1+r} \end{cases} \\ E[Y_{(O)}^{(L|P)}] &= \frac{\alpha(1+r) \left(E[X \wedge \frac{u}{1+r}] - E[X \wedge \frac{d}{1+r}] \right)}{S_X\left(\frac{d}{1+r}\right)} \end{aligned}$$

14 Estimation non-paramétrique des fonctions de répartition et de survie

Distribution empirique avec données complètes

$$\text{I.C. au niveau } 1 - \alpha \text{ de } F(x) \in \left[F_n(x) \pm z_{1-\frac{\alpha}{2}} \sqrt{\widehat{V}(F_n(x))} \right]$$

y_j : la j -ème des k valeurs unique de l'échantillon de n ($k \leq n$).
 $y_1 < y_2 < \dots < y_k$

s_j : Nombre de fois que l'observation y_j est observé dans l'échantillon.
 $\sum_{j=1}^k s_j = n$

r_j : Nombre d'observations $\geq y_j$.
 $\sum_{i=j}^k s_i = r_j$

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n I_{\{x_j \leq x\}}$$

$$f_n(x) = \frac{1}{n} \sum_{j=1}^n I_{\{x_j = x\}}$$

$$nF_n(x) \sim \text{bin}(n, F(x))$$

$$E[F_n(x)] = F_n(x)$$

$$\widehat{Var}[F_n(x)] = \frac{F_n(x)(1 - F_n(x))}{n}$$

$$F_n(x) = \begin{cases} 0, & x < y_1 \\ 1 - \frac{r_j}{n}, & y_{j-1} \leq x < y_j \\ 1, & x > y_k \end{cases}$$

$\forall j = 2, \dots, k$

Distribution empirique avec données groupées

Fonction OGIVE

- > Dans certains contextes, on a n données qui sont groupées en intervalles et la fonction OGIVE permet d'interpoler entre 2 points c_{j-1} et c_j .
- > On définit n_j comme étant le nombre d'observations entre c_{j-1} et c_j .
- > Soit x tel que

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

Alors

$$\begin{aligned} F_n^{\text{OGIVE}}(x) &= \alpha F_n(c_{j-1}) + (1 - \alpha) F_n(c_j) \\ &= \frac{c_j - x}{c_j - c_{j-1}} F_n(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}} F_n(c_j) \end{aligned}$$

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

$$f_n(x) = \frac{F_n(c_j) - F_n(c_{j-1})}{c_j - c_{j-1}} \Leftrightarrow \frac{n_j}{n(c_j - c_{j-1})}$$

Estimations empirique avec données censurées à droite

On représente les données censurées avec :

b_i : Nombre d'observations censurées à la droite dans l'intervalle $[y_i, y_{i+1}) \forall i = 1, 2, \dots, k-1$

De plus, on interprète les valeurs définies plus haut.

s_i : Nombre de décès au temps i .

r_i : Le nombre à *risque* à l'observation y_i .

$$r_i = \begin{cases} n, & i = 1 \\ r_{i-1} - s_{i-1} - b_{i-1}, & i = 2, 3, \dots, k+1 \end{cases}$$

On peut interpréter la fonction de survie comme une **probabilité conditionnelle**.

$$S(t) = \frac{S(t_1)}{S(t_0)} \times \frac{S(t_2)}{S(t_1)} \times \dots \times \frac{S(t)}{S(t-1)}$$

$$\Leftrightarrow p_1 \times p_2 \times \dots \times p_t = \prod_{j \leq t} p_j$$

où $p_t = P(T > t | T > t-1)$

On peut donc estimer p_j par :

$$\hat{p}_j = 1 - \frac{S_j}{r_j}$$

où $S_j \sim \text{Bin}(r_j, q_j)$

Ceci correspond donc à l'estimateur de **Kaplan-Meier** :

Estimateur Kaplan-Meier de la fonction de survie empirique

$$S_m(t) = \prod_{j \leq t} \left(1 - \frac{S_j}{r_j} \right)$$

La **Formule de Greenwood** estime la variance de la fonction de survie **Kaplan-Meier** :

Formule de Greenwood

$$\hat{V}(S_m(t)) = (S_m(t))^2 \sum_{j \leq t} \frac{S_j}{r_j(r_j - S_j)}$$

La **cumulative hazard rate function** est estimée par l'estimateur **Nelson-Åalen** :

Estimateur Nelson-Åalen du cumulative hazard rate function

$$\hat{H}(t) = \sum_{j \leq t} \frac{S_j}{r_j}$$

L'estimateur de la fonction de survie peut ensuite être déduit :

$$\hat{S}(t) = e^{-\hat{H}(t)}$$

La variance de l'estimateur Nelson-Åalen est estimée par la **formule de Klein** :

Formule de Klein

$$\hat{V}(\hat{H}(t)) = \sum_{j \leq t} \frac{S_j(r_j - S_j)}{r_j^3}$$

Intervalle de confiance

$$H(t) \in \left[\hat{H}(t) \pm z_{\alpha/2} \sqrt{\hat{V}(\hat{H}(t))} \right]$$

$$S(t) \in \left[S_m(t) \pm z_{\alpha/2} \sqrt{\hat{V}(S_m(t))} \right]$$

Méthode d'efron : Poser que $S_m(y) = 0 \forall y \geq y_{\max}$.

15 Fonction génératrice cumulée

Soit la fonction génératrice des moments $M_X(t)$, telle que

$$M_X(t) = E[e^{tX}]$$

Alors, la fonction génératrice cumulée $K_X(t)$ est définie comme

$$K(t) = \frac{\partial}{\partial t} \ln M_X(t)$$

De plus, la fonction génératrice cumulée a les propriétés suivantes :

$$K'(t) \Big|_{t=0} = E[X]$$

$$K''(t) \Big|_{t=0} = \text{Var}(X)$$

16 Frequentist estimation

Méthode des moments

On résout p équations à p inconnus, telles que

$$\hat{\mu}'_k = \mu'_k$$

Méthode des percentiles

On résout p équations à p inconnus (paramètres) telles que

$$F_n(\hat{\pi}_{g_i}) = g_i \quad i = 1, \dots, p$$

où $\hat{\pi}_{g_i}$ est le g_i quantile de la fonction empirique.

Smoothed empirical estimate

Parfois, le quantile recherché tombe entre 2 marches de la fonction empirique. On utilise l'approximation linéaire suivante avec les statistiques d'ordre $X_{(j)}$:

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

avec $j = \lfloor (n+1)g \rfloor$ et $h = (n+1)g - j$.

Méthode du maximum de vraisemblance (MLE)

Données complètes

On définit la fonction de vraisemblance $L(\theta)$ telle que

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

Et la fonction de log-vraisemblance $\ell(\theta)$

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

L'estimateur du maximum de vraisemblance θ maximiser $L(\theta)$ ou $\ell(\theta)$, i.e.

$$\left. \frac{\partial}{\partial \theta} \ell(\theta) \right|_{\theta=\hat{\theta}_{MLE}} = 0$$

Données groupées

Si les données sont groupées, alors on utilise une forme plus générale de la fonction de vraisemblance :

$$L(\theta) = \prod_{j=1}^k \left(F_X(c_j; \theta) - F_X(c_{j-1}; \theta) \right)^{n_j}$$

où F_X est la fonction de répartition théorique de la distribution qu'on suppose la distribution de notre estimateur MLE . Si les données sont censurées à la classe c_{j-1} , alors on utilise $(1 - F_X(c_{j-1}; \theta))$.

Variance des estimateurs et intervalle de confiance

Estimation de la variance de $\hat{\theta}$

L'information de Fisher $I(\theta)$ est définie par

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

Si l'information n'est pas connue, on peut l'estimer avec l'information observée :

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

Ainsi, on peut calculer la variance de l'estimateur $\hat{\theta}_{MLE}$ telle que

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

Intervalle de confiance pour $\hat{\theta}$

Lorsque $n \rightarrow \infty$, $\hat{\theta} \sim N(\theta, \text{Var}(\hat{\theta}))$. Alors, on peut trouver un IC pour l'estimateur au seuil $1 - \alpha$:

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

Méthode delta pour estimer la variance d'une transformation de $\hat{\theta}$

Lorsqu'on veut calculer la variance d'une autre quantité que le paramètre $\hat{\theta}$ lui-même, on peut utiliser la méthode Delta :

$$\text{Var}(h(\hat{\theta})) = \left(\frac{\partial}{\partial \theta} h(\theta) \right)^2 \text{Var}(\hat{\theta})$$

Dans un contexte multivarié, où $\hat{\theta}$ est un vecteur d'estimateurs, alors 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 $h(\theta)$:

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

Test du rapport de vraisemblance (LRT)

On veut tester si le modèle réduit avec θ_0 , qui est une bonne simplification de θ_1 , le modèle complet. Alors, on teste si la différence dans les log-vraisemblance est significative :

$$T = 2(\ell(\hat{\theta}) - \ell(\theta_0)) \sim \chi_{dl_1 - dl_0, 1-\alpha}^2$$

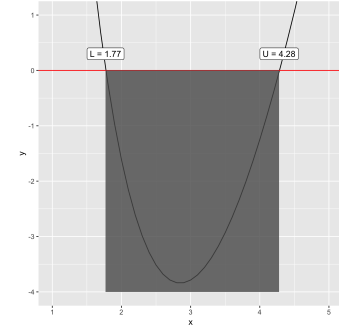
où dl_1 est le nombre de paramètres non-fixés du modèle complet et dl_0 le nombre de paramètres non-fixés du modèle réduit. **On va rejeter H_0 si $T > \chi_{dl_1 - dl_0, 1-\alpha}^2$ (test unilatéral)**, en concluant que le modèle réduit n'est pas une bonne simplification du modèle de l'hypothèse alternative.

Construction d'un intervalle de confiance par inversion du LRT

Si θ_0 est un paramètre adéquat pour le modèle réduit, alors la statistique T du LRT ne dépassera pas le quantile théorique $\chi_{dl_1 - dl_0, 1-\alpha}^2$. Alors, on veut trouver $\hat{\theta}_0$ tel que

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

On trouvera une équation du genre $g(\theta) \leq 0$, où g sera une fonction avec deux racines définies, qui correspondent aux bornes de l'intervalle de confiance pour les valeurs de $\hat{\theta}_0$:



17 Sélection de modèles

Chi-Square Goodness-of-fit

On veut valider l'adéquation du modèle qu'on propose avec ce test. On calcule la quantité X^2 :

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

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. On peut prouver que

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

On peut aussi faire le test LRT pour valider l'adéquation aussi.

Critères de sélection

Pour choisir entre plusieurs modèles, on peut, entre autres, se baser sur les critères suivants :

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**.

18 Estimation bayésienne

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 conditionnée par rapport à la valeur que θ prend (i.e. $f_{X|\Theta}$).

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 $\pi_{\Theta|X}(\theta|x)$:

$$\pi_{\Theta|X}(\theta|x) = \frac{f_{\Theta,X}(\theta, x)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta)\pi(\theta)}{\int f_{X|\Theta}(x|\theta)\pi(\theta)d\theta} \quad (1)$$

L'idée est de remplacer les différentes distributions dans l'Équation 1, 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* :

$$\hat{\theta}_{BAYES} = E[\Theta|X] \quad (2)$$

Certaines lois à savoir

| Loi | $\Pr(X = x)$ ou $f_X(x)$ | $E[X]$ | $Var(X)$ | $M_X(t)$ |
|--------------------------|---|--------------------------|----------------------------|---|
| $Bin(n, p)$ | $\binom{n}{x} p^x (1-p)^{n-x}$ | np | $np(1-p)$ | $((1-p) + p^t)^n$ |
| $Pois(\lambda)$ | $\frac{e^{-\lambda} \lambda^x}{x!}$ | λ | λ | $e^{\lambda(t-1)}$ |
| $Gamma(\alpha, \lambda)$ | $\frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$ | $\frac{\alpha}{\lambda}$ | $\frac{\alpha}{\lambda^2}$ | $\left(\frac{\lambda}{\lambda-t}\right)^\alpha$ |
| $Normale(\mu, \sigma^2)$ | $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ | μ | σ^2 | $e^{\mu t + \frac{\sigma^2 t^2}{2}}$ |

Rappels d'algèbre linéaire

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.

19 Rappel de probabilité