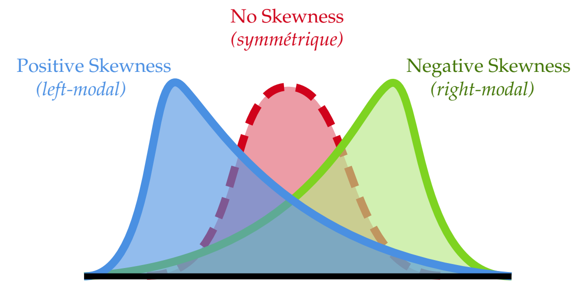
3 Estimation non-paramétrique

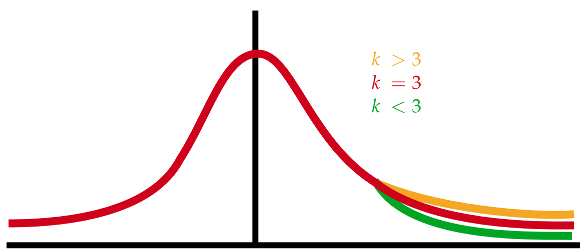
Moments à savoir

$$\mu'_k = \mathrm{E}\left[X^k\right] \qquad \mu_k = \mathrm{E}\left[(X-\mu)^k\right]$$

$$CV = \frac{\sigma}{\mu}$$

$$\text{Skewness : } \gamma = \frac{\mu_3}{\sigma^3} \qquad \text{Kurtosis : } \kappa = \frac{\mu_4}{\sigma^4}$$





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.

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

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.

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

Quantités des distributions à connaître

Y^P : **Excess loss**, alias **left truncated** and **shifted** variable. On interprete 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 égale à 0.

right censored : Toutes valeurs supérieures à u sont égale 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

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

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

$$\mathrm{E}\left[Y\right] = E[X\wedge d] = \int_0^d f_X(x)dx \Leftrightarrow \int_0^d S_X(x)dx$$

8 Fréquence et sévérité avec modifications aux contrats

Déductible ordinaire

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

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

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

Déductible franchise

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

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

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

Moments

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

Où (L | P) et (O | F) est à être interprété en REGEX.
C'est soit per loss (L) ou per payment (P)
C'est soit un déductible ordinaire (O) ou avec franchise (F).
De plus, on note que :

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

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

Fonctions

$$f_{Y_{(O)}^{(L|P)}} = \frac{f_X(y+d)}{S_X(d)}$$
$$S_{Y_{(O)}^{(L|P)}} = \frac{S_X(y+d)}{S_X(d)}$$
$$F_{Y_{(O)}^{(L|P)}} = \frac{F_X(y+d) - F_X(d)}{S_X(d)}$$
$$h_{Y_{(O)}^{(Y|P)}} = h_X(y+d)$$

LER et inflation du déductible ordinaire

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

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

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

$$Y = (X \wedge u) = \begin{cases} X, & X < u \\ u, & X \geq u \end{cases}$$
$$f_Y(y) = \begin{cases} f_X(y), & y < u \\ S_X(u), & y = u \end{cases}$$
$$F_Y(y) = \begin{cases} F_X(y), & y \leq u \\ 1, & y > u \end{cases}$$

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.

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

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

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

Distribution empirique avec données complètes

I.C. au niveau $1-\alpha$ 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 ($k \leq n$).
$y_1 < y_2 < \cdots < y_k$

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

$$F_n(x) = \frac{1}{n}\sum_{j=1}^n I_{\{x_j \leq x\}}$$
$$f_n(x) = \frac{1}{n}\sum_{j=1}^n I_{\{x_j = x\}}$$
$$nF_n(x) \sim \text{bin}(n, F(x))$$
$$E[F_n(x)] = \frac{nF_n(x)}{n} = F_n(x)$$
$$\widehat{Var}[F_n(x)] = \frac{nF_n(x)(1-F)n(x)}{n^2} = \frac{F_n(x)(1-F_n(x))}{n}$$
$$\widehat{Var}[S_n(x)] = \frac{S_n(x)(1-S_n(x))}{n}$$
$$F_n(x) = \begin{cases} 0, & x < y_1 \\ 1 - \frac{r_j}{n}, & y_{j-1} \leq x < y_j, j = 2, ..., k \\ 1, & x > y_k \end{cases}$$

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 défini n_j comme étant le nombre d'observations entre c_{j-1} et c_j.
- Soit x tel que

$$c_{j-1} \leq x \leq c_j$$
$$F_n(c_{j-1}) \leq F_n(x) \leq F_n(c_j)$$

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.

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

Par la suite, on peut interpréter la fonction de survie comme une probabilité conditionelle puisque $S(t_0) = 1$.

$$S(t) = \frac{S(t_1)}{S(t_0)} \times \frac{S(t_2)}{S(t_1)} \times \cdots \times \frac{S(t)}{S(t-1)}$$
$$\Leftrightarrow p_1 \times p_2 \times \cdots \times p_t = \prod_{j \leq t} p_j$$

$$\text{où } p_t = P(T > t | T > t-1)$$
$$\text{et } q_t = P(T < t | T > t-1)$$

On peut donc estimer p_j par :

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

$$\text{où } S_j \sim \text{Bin}(r_j, q_j)$$

Pour obtenir l'estimateur **Kaplan-Meier** :

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

On utilise la **méthode Delta** pour trouver la variance :

$$V(g(\hat{\theta})) = (g'(\hat{\theta}_0))^2 V(\hat{\theta})$$

Pour ensuite obtenir :

$$V(S_n(t)) = (S_n(t))^2 \sum_{j \leq t} \frac{q_j}{r_j p_j}$$

On peut donc estimer cette variance avec la **(Formule de Greenwood)** :

$$\widehat{V}[S_n(t)] = (S_n(t))^2 \sum_{j \leq t} \frac{S_j}{r_j(r_j - S_j)}$$

On reprends la même notation pour trouver l'estimateur **Nelson-Aalen** :

$$\widehat{H}(t) = \sum_{j \leq t} \left(\frac{S_j}{r_j}\right)$$

On peut ensuite déduire la fonction de survie empirique :

$$S_m(t) = e^{-\widehat{H}(t)}$$

La variance de la ***cumulative hazard rate function*** est représentée comme suit :

$$V(\widehat{H}(t)) = \sum_{j \leq t} V\left(\frac{S_j}{r_j}\right) = \frac{q_j p_j}{r_j}$$

Ce qu'on peut ensuite estimer avec la **formule de Klein** :

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

15 Fonction génératrice cumulante

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

$$M_X(t) = \mathrm{E}\left[e^{tX}\right]$$

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

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

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

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

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

16 Frequentist estimation

Méthode des moments

On résoud p équations à p inconnus, telles que

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

Méthode des percentiles

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

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

où $\hat{\pi}_{g_i}$ est le $g_i{}^e$ 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 $\boldsymbol{\theta}$ maximiser $L(\theta)$ ou $\ell(\theta)$, i.e.

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

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 censurés 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

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

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

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

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

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

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

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

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

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 :

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

Dans un contexte multivarié, où $\hat{\boldsymbol{\theta}}$ est un vecteur d'estimateurs, alors on a

$$\mathrm{Var}\,(h(\hat{\theta})) = \boldsymbol{h}^\top I(\theta)^{-1} \boldsymbol{h}$$

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

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

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 :

$$T = 2\left(\ell(\theta) - \ell(\theta_0)\right) \sim \chi^2_{dl_1 - dl_0, 1-\alpha}$$

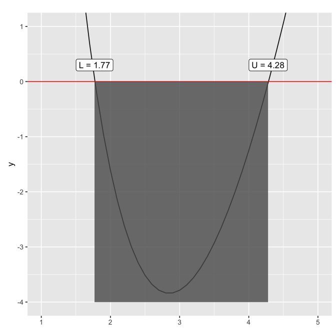
où dl_1 est le nombre de paramètres non-fixés du modèle complet et dl_0 le nombre de paramètres non-fixés du modèle réduit. **On va rejeter H_0 si $T > \chi^2_{dl_1 - dl_0, 1-\alpha}$ (test unilatéral)**, en concluant que le modèle réduit n'est pas une bonne simplification du modèle de l'hypothèse alternative.

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 $\chi_{dl, 1-\alpha^2}$. Alors, on veut trouver $\hat{\theta}_0$ tel que

$$2\left(\ell(\theta) - \ell(\theta_0)\right) \leq \chi^2_{dl_1 - dl_0, 1-\alpha}$$

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

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

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

Critères de sélection

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

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

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* :

$$\hat{\theta}_{BAYES} = \mathrm{E}\left[\Theta|X\right] \tag{2}$$

19 Rappel de probabilité

Certaines lois à savoir

Loi	$\Pr(X=x)$ ou $f_X(x)$	$\mathrm{E}[X]$	$Var(X)$	$M_X(t)$
$Bin(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	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 $\boldsymbol{A}^\top$, telle que

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

Déterminant d'une matrice

On peut calculer le déterminant $\det(\boldsymbol{A})$ de la matrice $\boldsymbol{A}$ tel que

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

Inverse d'une matrice

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

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

où on multiple par la matrice adjointe de $\boldsymbol{A}$. Il faut normalement calculer les cofacteurs, mais le cas à 2 dimensions est un cas simplifié.