



PRÉPASIMPLE

prepasimple.fr

HEC / ESCP

Mathématiques 2

1995

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Partie I. La loi de Pareto

Les paramètres vérifient $\alpha > 0$ et $x_0 + C > 0$. Dans la loi à deux paramètres, $C = 0$ et $x_0 > 0$.

A. Quelques résultats probabilistes

1.a)

Pour $x < x_0$, $f(x) = 0$. Pour $x \geq x_0$, on a $x + C \geq x_0 + C > 0$, donc la puissance du dénominateur est définie et $f(x) > 0$.

La fonction est continue sur chacun des intervalles $] -\infty, x_0[$ et $]x_0, +\infty[$; une éventuelle discontinuité en x_0 est compatible avec la définition d'une densité.

Pour $R > x_0$, une primitive donne

$$\int_{x_0}^R f(x) \, dx = \left[-\frac{(x_0 + C)^\alpha}{(x + C)^\alpha} \right]_{x_0}^R = 1 - \left(\frac{x_0 + C}{R + C} \right)^\alpha.$$

Comme $\alpha > 0$, le dernier terme tend vers 0 lorsque $R \rightarrow +\infty$. L'intégrale impropre converge donc et vaut 1.

La positivité, la régularité requise et la normalisation sont vérifiées : f est une densité.

1.b)

Notons F_X la fonction de répartition. Pour $x < x_0$, $X \geq x_0$ presque sûrement, donc $F_X(x) = 0$. Pour $x \geq x_0$, le calcul précédent donne

$$F_X(x) = 1 - \left(\frac{x_0 + C}{x + C} \right)^\alpha.$$

Ainsi

$$F_X(x) = \begin{cases} 0 & \text{si } x < x_0, \\ 1 - \left(\frac{x_0 + C}{x + C} \right)^\alpha & \text{si } x \geq x_0. \end{cases}$$

En particulier, $F_X(x_0) = 0$. La variable ayant une densité, $P(X = x_0) = 0$; les versions strictes et larges des probabilités de queue coïncident.

2.a)

La variable X est positive. Elle admet une espérance si et seulement si l'intégrale de transfert suivante converge :

$$\int_{x_0}^{+\infty} x f(x) \, dx = \alpha x_0^\alpha \int_{x_0}^{+\infty} x^{-\alpha} \, dx.$$

Il s'agit d'une intégrale de puissance à l'infini, qui converge exactement lorsque $\alpha > 1$.

Pour $\alpha > 1$, sur un segment $[x_0, R]$ puis par passage à la limite,

$$\int_{x_0}^{+\infty} x^{-\alpha} \, dx = \frac{x_0^{1-\alpha}}{\alpha - 1}.$$

On obtient

$$E(X) = \frac{\alpha x_0}{\alpha - 1} \quad \text{si } \alpha > 1.$$

Pour $\alpha = 1$, l'intégrale diverge comme $\ln R$; pour $0 < \alpha < 1$, elle diverge comme $R^{1-\alpha}$. Dans ces deux cas, X n'admet pas d'espérance finie.

2.b)

L'existence d'une variance nécessite un moment d'ordre 2 fini. Comme $X \geq 0$, on étudie

$$\int_{x_0}^{+\infty} x^2 f(x) \, dx = \alpha x_0^\alpha \int_{x_0}^{+\infty} x^{1-\alpha} \, dx.$$

Cette intégrale converge exactement lorsque $\alpha > 2$. Cette condition implique $\alpha > 1$, donc assure aussi l'existence de l'espérance.

Pour $\alpha > 2$,

$$E(X^2) = \frac{\alpha x_0^2}{\alpha - 2}.$$

Par la formule de Huygens,

$$V(X) = \frac{\alpha x_0^2}{\alpha - 2} - \frac{\alpha^2 x_0^2}{(\alpha - 1)^2}.$$

En mettant au même dénominateur, le numérateur restant est

$$\alpha((\alpha - 1)^2 - \alpha(\alpha - 2)) = \alpha.$$

Ainsi

$$V(X) = \frac{\alpha x_0^2}{(\alpha - 1)^2(\alpha - 2)} \quad \text{si } \alpha > 2.$$

Pour $\alpha = 2$, le moment d'ordre 2 diverge logarithmiquement ; pour $0 < \alpha < 2$, il diverge également. Il n'existe donc pas de variance finie dans ces cas.

3.

Comme $\lambda > 0$, l'événement $\{\lambda X \leq y\}$ équivaut à $\{X \leq y/\lambda\}$. Le seuil inférieur devient $\lambda x_0 > 0$.

Si $y < \lambda x_0$, $F_Y(y) = 0$. Si $y \geq \lambda x_0$,

$$F_Y(y) = F_X(y/\lambda) = 1 - \left(\frac{x_0}{y/\lambda} \right)^\alpha = 1 - \left(\frac{\lambda x_0}{y} \right)^\alpha.$$

On reconnaît la fonction de répartition d'une loi de Pareto à deux paramètres :

$$Y \sim VP(\alpha, \lambda x_0).$$

Une multiplication par un nombre strictement positif conserve le paramètre de forme α .

4.a)

La variable $U = X + \mu$ est à valeurs dans $[x_0 + \mu, +\infty[$. Pour tout réel u ,

$$F_U(u) = F_X(u - \mu).$$

Pour $u \geq x_0 + \mu$,

$$F_U(u) = 1 - \left(\frac{x_0}{u - \mu} \right)^\alpha.$$

Cette expression est celle d'une loi de Pareto à trois paramètres, avec seuil $x_0 + \mu$ et paramètre de translation $-\mu$. La condition d'admissibilité vaut

$$(x_0 + \mu) + (-\mu) = x_0 > 0.$$

Ainsi

$$U \sim VP(\alpha, x_0 + \mu, -\mu).$$

Cette conclusion reste valable même si $x_0 + \mu$ est négatif : seule la somme du seuil et du paramètre de translation doit être strictement positive.

4.b)

La variable $V = Z + C$ est à valeurs dans $[x_0 + C, +\infty[$, avec $x_0 + C > 0$.

Pour $v \geq x_0 + C$,

$$F_V(v) = F_Z(v - C) = 1 - \left(\frac{x_0 + C}{v} \right)^\alpha.$$

En dessous de ce seuil, la fonction vaut 0. Ainsi

$$V \sim VP(\alpha, x_0 + C).$$

4.c)

Posons $b = x_0 + C > 0$. D'après 4.b), $V = Z + C$ suit $VP(\alpha, b)$.

L'intégrabilité de Z équivaut à celle de V , car

$$|Z| \leq V + |C|, \quad V \leq |Z| + |C|.$$

D'après 2.a), elle a donc lieu exactement lorsque $\alpha > 1$. Dans ce cas,

$$E(Z) = E(V) - C = \frac{\alpha(x_0 + C)}{\alpha - 1} - C = \frac{\alpha x_0 + C}{\alpha - 1}.$$

De même, les inégalités $Z^2 \leq 2V^2 + 2C^2$ et $V^2 \leq 2Z^2 + 2C^2$ montrent que les moments d'ordre 2 existent simultanément. D'après 2.b), cela a lieu exactement pour $\alpha > 2$.

Une translation ne change pas la variance. Pour $\alpha > 2$,

$$V(Z) = V(V) = \frac{\alpha(x_0 + C)^2}{(\alpha - 1)^2(\alpha - 2)}.$$

5.

Comme $W \geq 0$ presque sûrement et $k > 1$, on a $T = x_0 k^W \geq x_0$. Pour $t < x_0$, $F_T(t) = 0$.

Pour $t \geq x_0$, la fonction $w \mapsto k^w$ est strictement croissante et $\ln k > 0$, donc

$$\{T \leq t\} = \left\{ W \leq \frac{\ln(t/x_0)}{\ln k} \right\}.$$

La borne de droite est positive ou nulle. La fonction de répartition de la loi exponentielle donne

$$F_T(t) = 1 - \exp\left(-\frac{\beta}{\ln k} \ln(t/x_0)\right) = 1 - \left(\frac{x_0}{t}\right)^{\beta/\ln k}.$$

Puisque $\beta/\ln k > 0$, on reconnaît

$$T \sim VP\left(\frac{\beta}{\ln k}, x_0\right).$$

La transformation exponentielle d'une variable exponentielle fournit donc une loi de Pareto.

6.

Notons $R = \sqrt{X}$. Comme $X \geq x_0 > 0$, $R \geq \sqrt{x_0}$.

Pour $r < \sqrt{x_0}$, $F_R(r) = 0$. Pour $r \geq \sqrt{x_0}$, le réel r est positif et

$$F_R(r) = P(X \leq r^2) = 1 - \left(\frac{x_0}{r^2}\right)^\alpha = 1 - \left(\frac{\sqrt{x_0}}{r}\right)^{2\alpha}.$$

Ainsi

$$\sqrt{X} \sim VP(2\alpha, \sqrt{x_0}).$$

B. Propriété caractéristique de la loi de Pareto

1.

L'espérance de X existe, donc $\int_{\mathbb{R}} |t|f(t) dt < +\infty$. Le numérateur définissant $M_X(x)$ existe absolument. Le dénominateur est strictement positif par hypothèse.

On écrit

$$M_X(x) - x = \frac{\int_x^{+\infty} (t - x)f(t) dt}{\int_x^{+\infty} f(t) dt}.$$

Le numérateur est positif ou nul. Pour établir sa stricte positivité, on utilise le fait qu'une variable à densité n'a pas d'atome :

$$P(X > x) = P(X \geq x) = \int_x^{+\infty} f(t) dt > 0.$$

Les événements $\{X \geq x + 1/m\}$ augmentent vers $\{X > x\}$ lorsque l'entier m tend vers l'infini. Par continuité monotone des probabilités, il existe $m \geq 1$ tel que $P(X \geq x + 1/m) > 0$.

Sur $[x + 1/m, +\infty[$, $t - x \geq 1/m$. Ainsi

$$\int_x^{+\infty} (t - x)f(t) dt \geq \frac{1}{m}P(X \geq x + 1/m) > 0.$$

Le quotient est donc strictement positif, d'où $M_X(x) > x$.

2.

Méthode 1 — Quotient des deux intégrales de queue

Pour $x \geq x_0$, la queue de la loi de Pareto vaut

$$\int_x^{+\infty} f(t) dt = \left(\frac{x_0}{x}\right)^\alpha.$$

Comme $\alpha > 1$, le numérateur converge et vaut

$$\int_x^{+\infty} t f(t) dt = \alpha x_0^\alpha \int_x^{+\infty} t^{-\alpha} dt = \frac{\alpha x_0^\alpha x^{1-\alpha}}{\alpha - 1}.$$

En divisant par la queue, qui est strictement positive,

$$M_X(x) = \frac{\alpha}{\alpha - 1} x.$$

Méthode 2 — Loi conditionnelle au-dessus d'un seuil

Fixons $x \geq x_0$. L'événement $A = \{X \geq x\}$ a une probabilité strictement positive. Pour $t \geq x$,

$$P(X \geq t \mid A) = \frac{P(X \geq t)}{P(X \geq x)} = \frac{(x_0/t)^\alpha}{(x_0/x)^\alpha} = \left(\frac{x}{t}\right)^\alpha.$$

La loi conditionnelle de X sachant A est donc $VP(\alpha, x)$. Son espérance existe, car $\alpha > 1$, et vaut $\alpha x / (\alpha - 1)$ par I.A.2.a).

L'espérance conditionnelle à cet événement est précisément le quotient d'intégrales définissant $M_X(x)$. On retrouve la même formule. Cette seconde méthode montre que relever le seuil d'observation conserve le paramètre de forme de la loi.

3.a)

La densité est continue sur $[x_0, +\infty[$, et X est intégrable. Comme $X \geq x_0 > 0$ presque sûrement,

$$G(x_0) = 1, \quad H(x_0) = E(X).$$

Pour $x \geq x_0$, la relation de Chasles pour les intégrales convergentes donne

$$G(x) = 1 - \int_{x_0}^x f(t) \, dt, \quad H(x) = E(X) - \int_{x_0}^x t f(t) \, dt.$$

Les intégrandes sont continues sur tout segment de $[x_0, +\infty[$. Le théorème fondamental du calcul intégral donne donc

$$G'(x) = -f(x), \quad H'(x) = -x f(x).$$

Ces formules valent sur $]x_0, +\infty[$ et au sens de la dérivée à droite en x_0 . On n'a pas dérivé une intégrale impropre sous son signe : on l'a d'abord écrite comme une constante moins une intégrale sur un segment.

3.b)

Pour tout $x \geq x_0$, la stricte positivité et la continuité de f assurent $G(x) > 0$. L'hypothèse $M_X(x) = kx$ s'écrit donc

$$H(x) = kxG(x).$$

Les fonctions étant dérivables, on peut dériver cette identité. Par 3.a),

$$xG'(x) = kG(x) + kxG'(x).$$

En regroupant,

$$kG(x) = (1 - k)xG'(x).$$

Puisque $k > 1$, on a en particulier $k \neq 0$, et

$$G(x) = \frac{1 - k}{k} xG'(x).$$

3.c)

Posons $q = k/(k - 1) > 1$.

Méthode 1 — Produit de dérivée nulle, proposé par l'énoncé

Comme $x \geq x_0 > 0$, la fonction $I(x) = x^q G(x)$ est dérivable. La relation de 3.b) équivaut à $xG'(x) = -qG(x)$, donc

$$I'(x) = qx^{q-1}G(x) + x^q G'(x) = x^{q-1}(qG(x) + xG'(x)) = 0.$$

La fonction I est constante sur l'intervalle $[x_0, +\infty[$, par la caractérisation des fonctions de dérivée nulle. En x_0 , elle vaut $x_0^q G(x_0) = x_0^q$. Ainsi

$$G(x) = \left(\frac{x_0}{x}\right)^q \quad (x \geq x_0).$$

Méthode 2 — Dérivée logarithmique

La fonction G est strictement positive. On peut donc diviser par $G(x)$ et écrire

$$\frac{G'(x)}{G(x)} = -\frac{q}{x}.$$

Les deux membres sont continus sur tout segment $[x_0, x]$. En intégrant,

$$\ln G(x) - \ln G(x_0) = -q(\ln x - \ln x_0).$$

Comme $G(x_0) = 1$, l'exponentiation donne $G(x) = (x_0/x)^q$.

La première méthode suit la fonction auxiliaire imposée par le sujet ; la seconde utilise la positivité de la queue pour obtenir la même expression par le logarithme.

3.d)

Pour $x < x_0$, $F_X(x) = 0$. Pour $x \geq x_0$, l'absence d'atome et 3.c) donnent

$$F_X(x) = 1 - G(x) = 1 - \left(\frac{x_0}{x}\right)^{k/(k-1)}.$$

C'est la fonction de répartition de la loi de Pareto à deux paramètres correspondante. Comme la fonction de répartition caractérise la loi,

$$X \sim VP\left(\frac{k}{k-1}, x_0\right).$$

Le paramètre obtenu est strictement supérieur à 1, ce qui est cohérent avec l'hypothèse d'existence de l'espérance.

4.a)

Une translation préserve l'intégrabilité : $|Y| \leq |X| + |\lambda|$. La variable $Y = X + \lambda$ admet la densité $u \mapsto f(u - \lambda)$.

Pour une valeur y où le dénominateur est non nul, le changement de variable affine $t = u - \lambda$, d'abord sur un segment puis dans les intégrales absolument convergentes, donne

$$\int_y^{+\infty} f(u - \lambda) \, du = \int_{y-\lambda}^{+\infty} f(t) \, dt$$

et

$$\int_y^{+\infty} u f(u - \lambda) \, du = \int_{y-\lambda}^{+\infty} t f(t) \, dt + \lambda \int_{y-\lambda}^{+\infty} f(t) \, dt.$$

En divisant la seconde identité par la première,

$$M_Y(y) = M_X(y - \lambda) + \lambda.$$

Les dénominateurs des deux moyennes sont égaux ; elles sont donc définies simultanément. Sous les hypothèses de B.4, les queues sont positives pour tout seuil fini.

4.b)

D'après I.A.4.b), $V = X + C$ suit $VP(\alpha, x_0 + C)$. Pour $x \geq x_0$, on a $x + C \geq x_0 + C > 0$, donc B.2 s'applique :

$$M_V(x + C) = \frac{\alpha}{\alpha - 1}(x + C).$$

D'après B.4.a), $M_V(x + C) = M_X(x) + C$. On en déduit

$$M_X(x) = \frac{\alpha}{\alpha - 1}x + \frac{C}{\alpha - 1}.$$

La moyenne au-dessus d'un seuil est donc une fonction affine de ce seuil. Sa pente est strictement supérieure à 1.

4.c)

Posons $c = h/(k-1)$ et $Y = X + c$. La variable Y est à valeurs dans $[x_0 + c, +\infty[$, avec $x_0 + c > 0$ par hypothèse. Sa densité est continue et strictement positive sur cet intervalle, et son espérance existe par translation.

Pour $y \geq x_0 + c$, B.4.a) et l'hypothèse donnent

$$M_Y(y) = M_X(y - c) + c = k(y - c) + h + c = ky + h - (k-1)c = ky.$$

Toutes les hypothèses de B.3 sont vérifiées pour Y avec seuil $x_0 + c$. On conclut

$$Y \sim VP\left(\frac{k}{k-1}, x_0 + c\right).$$

Enfin, $X = Y - c$. La translation décrite en I.A.4.a), avec $\mu = -c$, donne

$$X \sim VP\left(\frac{k}{k-1}, x_0, c\right) = VP\left(\frac{k}{k-1}, x_0, \frac{h}{k-1}\right).$$

La condition sur les paramètres est bien $x_0 + c > 0$.

5. Un exemple statistique : la répartition des revenus

5.1.a)

On lit la pente de D à partir de deux points éloignés de la droite, et non à partir de deux observations quelconques du nuage.

Le graphique original a des graduations espacées de 100 sur les deux axes. Après calibration de ces graduations, la droite monte d'environ 660 KF lorsque l'abscisse augmente de 400 KF. Sa pente est donc voisine de

$$k \simeq \frac{660}{400} = 1,65.$$

L'extrapolation vers l'axe des ordonnées donne une ordonnée à l'origine de l'ordre de 30 KF. Pour les calculs de modélisation suivants, on retient donc

$$D : M(x) \simeq 1,65x + 30.$$

La lecture de l'ordonnée à l'origine est moins précise que celle de la pente, car la droite dessinée ne rejoint pas exactement l'axe et les graduations sont espacées. Ces valeurs sont des approximations graphiques.

5.1.b)

La question I.B.4.c) a établi la caractérisation suivante : si une variable à densité continue et strictement positive sur un intervalle $[x_0, +\infty[$ possède une espérance et vérifie exactement $M_X(x) = kx + h$ avec $k > 1$ et $x_0 + h/(k - 1) > 0$, alors elle suit une loi de Pareto à trois paramètres.

Ici, le nuage suggère une moyenne de queue approximativement affine, avec une pente supérieure à 1. En retenant un modèle continu satisfaisant les hypothèses de cette caractérisation, et en prolongeant la relation affine sur son domaine, une loi de Pareto à trois paramètres constitue donc un modèle possible des revenus.

Les observations d'un nombre fini de points ne prouvent pas que la loi réelle soit exactement une loi de Pareto ; elles motivent la modélisation proposée.

5.1.c)

Pour une loi $VP(\alpha, x_0, C)$ d'espérance finie, I.B.4.b) donne

$$M_X(x) = \frac{\alpha}{\alpha - 1}x + \frac{C}{\alpha - 1}.$$

La comparaison avec $kx + h$ donne

$$k = \frac{\alpha}{\alpha - 1}, \quad h = \frac{C}{\alpha - 1}.$$

Comme $k > 1$, ces relations se résolvent en

$$\alpha = \frac{k}{k - 1}, \quad C = \frac{h}{k - 1}.$$

Avec $k \simeq 1,65$ et $h \simeq 30$,

$$\alpha \simeq \frac{1,65}{0,65} \simeq 2,54, \quad C \simeq \frac{30}{0,65} \simeq 46 \text{ KF}.$$

Ces estimations dépendent de la précision de lecture de D , en particulier de son ordonnée à l'origine.

5.1.d)

Dans le modèle, $X \geq x_0$ presque sûrement et X possède une densité. La moyenne au-dessus du seuil minimal est donc la moyenne de toute la population :

$$M_X(x_0) = E(X) = 75.$$

Avec le modèle affine retenu, $kx_0 + h \simeq 75$, d'où

$$x_0 \simeq \frac{75 - h}{k} \simeq \frac{75 - 30}{1,65} \simeq 27 \text{ KF}.$$

On vérifie $x_0 + C > 0$, puisque les deux valeurs estimées sont positives. Le modèle obtenu possède bien une espérance, puisque $\alpha \simeq 2,54 > 1$.

5.2.a)

Question adaptée au programme actuel — Le graphique 2 annoncé dans l'énoncé original n'est pas présent dans le PDF fourni. Dans cette version, ses neuf points sont calculés à partir du tableau avec $C_{\text{gr}} = 46$, et la droite Δ est explicitement définie par les points extrêmes $x = 40$ et $x = 500$. On ne prétend pas retrouver une droite de l'original qui n'est pas disponible.

Les deux points définissant Δ ont pour coordonnées

$$A = (\ln 86, \ln 658), \quad B = (\ln 546, \ln 6).$$

Les abscisses sont distinctes puisque $86 \neq 546$. Le coefficient directeur vaut donc

$$a = \frac{\ln 6 - \ln 658}{\ln 546 - \ln 86} = \frac{\ln(6/658)}{\ln(546/86)} \simeq -2,5415.$$

L'ordonnée à l'origine est

$$b = \ln 658 - a \ln 86 \simeq 17,8101.$$

La droite construite a donc pour équation approchée $v = -2,5415u + 17,8101$.

5.2.b)

Pour une loi de Pareto à trois paramètres et $x \geq x_0$, I.A.1.b) donne

$$1 - F(x) = \left(\frac{x_0 + C}{x + C} \right)^\alpha.$$

Tous les arguments des logarithmes sont strictement positifs. En prenant le logarithme,

$$\ln(1\,000(1 - F(x))) = -\alpha \ln(x + C) + \ln 1\,000 + \alpha \ln(x_0 + C).$$

Les coordonnées transformées sont donc exactement alignées dans un modèle de Pareto idéal. La pente est $-\alpha$ et l'ordonnée à l'origine $\ln 1\,000 + \alpha \ln(x_0 + C)$.

Dans le graphique reconstruit, les points sont approximativement alignés suivant une pente négative ; cette représentation est donc cohérente avec une queue de Pareto. Les écarts visibles à la droite et les arrondis du tableau ne permettent pas de conclure à une égalité exacte de lois.

5.2.c)

On identifie la pente de la droite construite à $-\alpha$, d'où

$$\alpha \simeq -a \simeq 2,5415.$$

L'identification de l'ordonnée à l'origine donne

$$b \simeq \ln 1\,000 + \alpha \ln(x_0 + C).$$

Comme $\alpha > 0$, on résout cette relation :

$$x_0 \simeq \exp\left(\frac{b - \ln 1\,000}{\alpha}\right) - C.$$

Avec $C \simeq 46$, $b \simeq 17,8101$ et $\alpha \simeq 2,5415$,

$$x_0 \simeq 26,94 \text{ KF} \simeq 27 \text{ KF}.$$

On retrouve des valeurs compatibles avec celles obtenues sur le graphique 1. Cette cohérence porte sur le modèle construit à partir des données disponibles, avec le choix de la droite précisé dans le sujet adapté.

Partie II. Courbe de concentration et inégalité des revenus

On note désormais $\mathcal{C}$ la fonction de concentration pour la distinguer du paramètre C d'une loi de Pareto. On pose $m = E(X)$ lorsque l'espérance existe dans le cadre étudié.

A. Courbe de concentration et indice de Gini

1.a)

La variable est à valeurs dans $[x_0, +\infty[$ avec $x_0 > 0$ et une densité continue strictement positive sur cet intervalle. Sur $]x_0, +\infty[$,

$$F'(x) = f(x) > 0.$$

La fonction F est donc continue et strictement croissante sur $[x_0, +\infty[$, avec $F(x_0) = 0$.

La normalisation de la densité donne $F(x) \rightarrow 1$ en $+\infty$.

Pour tout x fini, $F(x) < 1$, car l'intégrale de f sur $[x, x+1]$ est strictement positive.

L'image de la restriction de F est donc exactement $[0, 1[$.

Par le théorème de la bijection, cette restriction possède une application réciproque continue et strictement croissante

$$F^{-1} : [0, 1[\longrightarrow [x_0, +\infty[.$$

En particulier, $F^{-1}(0) = x_0$. Il n'existe aucune valeur réelle $F^{-1}(1)$.

1.b)

On a $m \geq x_0 > 0$, donc la division par m est permise. La fonction Q est continue sur $[x_0, +\infty[$, avec $Q(x_0) = 0$, et

$$Q'(x) = \frac{xf(x)}{m} > 0 \quad (x > x_0).$$

Elle est donc strictement croissante. Par le transfert et l'existence de l'espérance,

$$\lim_{x \rightarrow +\infty} Q(x) = \frac{1}{m} \int_{x_0}^{+\infty} x f(x) \, dx = 1.$$

Pour tout x fini, $Q(x) < 1$, car la queue de l'intégrale de $x f(x)$ est strictement positive.

Sur $[0, 1[$, la composition $\mathcal{C} = Q \circ F^{-1}$ est donc continue, strictement croissante, à valeurs dans $[0, 1[$, et $\mathcal{C}(0) = Q(x_0) = 0$.

Lorsque $t \rightarrow 1^-$, on a $F^{-1}(t) \rightarrow +\infty$. En effet, pour tout $A > x_0$, $F(A) < 1$, et $t > F(A)$ implique $F^{-1}(t) > A$.

Ainsi $\mathcal{C}(t) \rightarrow 1$ lorsque $t \rightarrow 1^-$. On définit $\mathcal{C}(1) = 1$: le prolongement est continu sur $[0, 1]$. Il demeure strictement croissant, puisque $\mathcal{C}(t) < 1$ pour tout $t < 1$.

On a finalement

$$\mathcal{C}(0) = 0, \quad \mathcal{C}(1) = 1.$$

2.a)

Question adaptée au programme actuel — L'énoncé original demandait une dérivabilité sur $[0, 1]$ et utilisait $F^{-1}(1)$. Cette valeur n'existe pas et la pente devient infinie à l'extrémité 1. La formule est donc demandée sur $[0, 1[$, avec une dérivée à droite en 0.

Pour $t \in]0, 1[$, posons $x = F^{-1}(t) > x_0$. La fonction F est dérivable et sa dérivée $f(x)$ est strictement positive. Le théorème de dérivation de la réciproque donne

$$(F^{-1})'(t) = \frac{1}{f(F^{-1}(t))}.$$

Par la dérivation de la composition,

$$\mathcal{C}'(t) = Q'(F^{-1}(t))(F^{-1})'(t) = \frac{F^{-1}(t)f(F^{-1}(t))}{m} \times \frac{1}{f(F^{-1}(t))} = \frac{F^{-1}(t)}{m}.$$

Pour traiter $t = 0$, prenons $t > 0$ et $x = F^{-1}(t) > x_0$. On a

$$\frac{\mathcal{C}(t) - \mathcal{C}(0)}{t} = \frac{\int_{x_0}^x u f(u) \, du}{m \int_{x_0}^x f(u) \, du}.$$

Le dénominateur est strictement positif. Sur $[x_0, x]$, $x_0 \leq u \leq x$ et $f(u) \geq 0$, donc

$$\frac{x_0}{m} \leq \frac{\mathcal{C}(t)}{t} \leq \frac{x}{m}.$$

Lorsque $t \rightarrow 0^+$, la continuité de la réciproque donne $x \rightarrow x_0$. Par encadrement, la dérivée à droite existe et vaut $x_0/m = F^{-1}(0)/m$.

La formule annoncée est donc établie sur tout $[0, 1[$ avec la convention précisée au bord.

2.b)

La fonction F^{-1} est strictement croissante et $m > 0$. La formule de 2.a) montre, sans nouveau calcul de dérivée, que $\mathcal{C}'$ est strictement croissante sur $[0, 1[$.

La caractérisation de la convexité par la croissance de la dérivée donne la convexité de $\mathcal{C}$ sur cet intervalle. Son prolongement continu est convexe sur $[0, 1]$: les inégalités de convexité impliquant l'extrémité 1 s'obtiennent en passant à la limite depuis des points inférieurs à 1.

Le graphe d'une fonction convexe se situe sous toute corde joignant deux points de son graphe. La corde joignant $(0, 0)$ à $(1, 1)$ est la première bissectrice. Ainsi

$$\forall t \in [0, 1], \quad \mathcal{C}(t) \leq t.$$

On a suivi la consigne « sans calcul » en utilisant la monotonie de la réciproque et les propriétés géométriques de la convexité.

2.c)

D'après 2.a),

$$\mathcal{C}'(0) = \frac{x_0}{E(X)}.$$

La stricte positivité de la densité au-dessus de x_0 donne $E(X) > x_0$; on peut également appliquer I.B.1 au seuil x_0 , où $M_X(x_0) = E(X)$. Ainsi

$$0 < \mathcal{C}'(0) < 1.$$

D'après 1.b), $F^{-1}(t) \rightarrow +\infty$ lorsque $t \rightarrow 1^-$. L'espérance m est une constante finie strictement positive, donc

$$\lim_{t \rightarrow 1^-} \mathcal{C}'(t) = +\infty.$$

Cette limite exclut une dérivée réelle finie en $t = 1$.

2.d)

La courbe relie $(0, 0)$ à $(1, 1)$, est strictement croissante et convexe, et reste sous la première bissectrice.

En $(0, 0)$, sa tangente à droite a pour équation

$$y = \frac{x_0}{E(X)}t,$$

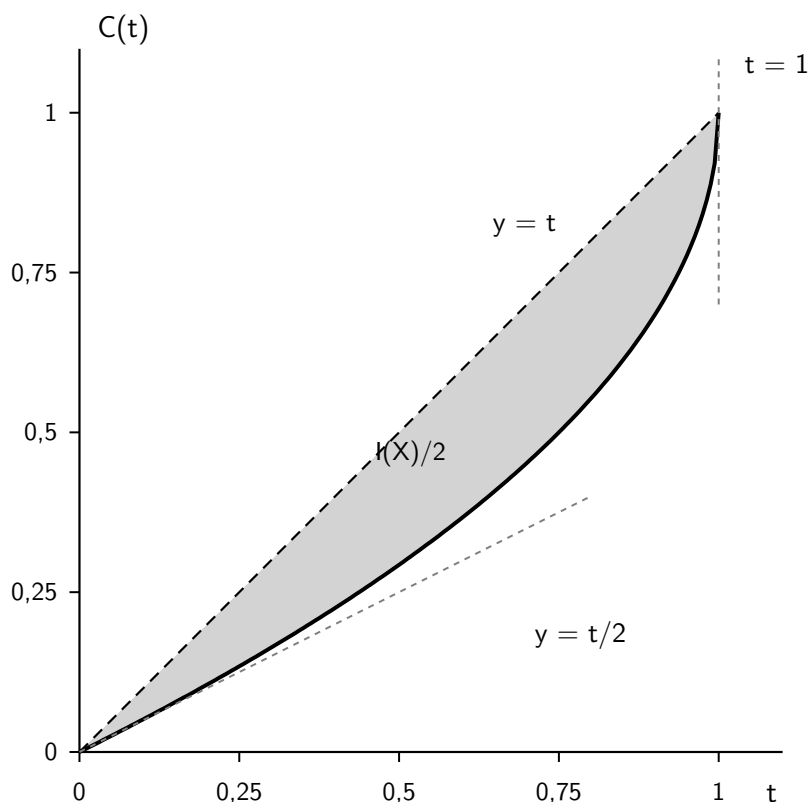
avec une pente comprise strictement entre 0 et 1.

Pour préciser la tangente en $(1, 1)$, appliquons le théorème des accroissements finis sur $[t, 1]$, avec $0 < t < 1$. La fonction est continue sur ce segment et dérivable sur son intérieur, donc il existe $\xi_t \in]t, 1[$ tel que

$$\frac{1 - \mathcal{C}(t)}{1 - t} = \mathcal{C}'(\xi_t).$$

Comme $\mathcal{C}'$ est croissante, le membre de droite est au moins $\mathcal{C}'(t)$, qui tend vers $+\infty$. La pente des sécantes tend donc vers $+\infty$: la tangente à gauche en $(1, 1)$ est verticale, d'équation $t = 1$.

Le dessin suivant illustre cette allure avec la courbe exacte $\mathcal{C}(t) = 1 - \sqrt{1 - t}$, correspondant à un exemple de Pareto avec $\alpha = 2$. Il ne fixe pas la loi dans la présente partie générale. La tangente à l'origine de cet exemple a pour pente $1/2$; dans le cas général, elle a la pente $x_0/E(X)$ indiquée ci-dessus.



Exemple : $\alpha = 2$

L'indice $I(X) = 2 \int_0^1 (t - \mathcal{C}(t)) dt$ existe, car l'intégrande est continue sur $[0, 1]$. La

convexité donne sa positivité ou nullité ; le dessin représente l'aire dont il est le double.

B. Comparaison de quelques procédures d'imposition des revenus

On suppose ici $X \sim VP(\alpha, x_0)$, avec $\alpha > 1$ et $x_0 > 0$. Son espérance vaut

$$m = E(X) = \frac{\alpha x_0}{\alpha - 1}.$$

1.a)

Pour $x \geq x_0$, on calcule sur le segment $[x_0, x]$:

$$\int_{x_0}^x t f(t) dt = \alpha x_0^\alpha \int_{x_0}^x t^{-\alpha} dt = \frac{\alpha x_0^\alpha}{\alpha - 1} (x_0^{1-\alpha} - x^{1-\alpha}).$$

En divisant par $m = \alpha x_0 / (\alpha - 1) > 0$,

$$Q(x) = 1 - \left(\frac{x_0}{x}\right)^{\alpha-1}.$$

On vérifie $Q(x_0) = 0$ et $Q(x) \rightarrow 1$ en $+\infty$.

1.b)

Méthode 1 — Composition avec la réciproque de la fonction de répartition

Pour $x \geq x_0$, $F(x) = 1 - (x_0/x)^\alpha$. Si $t \in [0, 1[$, la résolution de $F(x) = t$ donne

$$F^{-1}(t) = \frac{x_0}{(1-t)^{1/\alpha}}.$$

En substituant cette expression dans Q ,

$$\mathcal{C}(t) = Q(F^{-1}(t)) = 1 - (1 - t)^{(\alpha-1)/\alpha}.$$

L'exposant $(\alpha - 1)/\alpha$ est strictement positif. Le membre de droite tend donc vers 1 lorsque $t \rightarrow 1^-$, ce qui correspond au prolongement $\mathcal{C}(1) = 1$. La formule vaut ainsi sur tout $[0, 1]$.

Méthode 2 — Intégration de la dérivée de la courbe de concentration

La réciproque trouvée en résolvant $F(x) = t$ vaut $x_0(1 - t)^{-1/\alpha}$. La formule générale II.A.2.a) donne, pour $t < 1$,

$$\mathcal{C}'(t) = \frac{x_0(1 - t)^{-1/\alpha}}{\alpha x_0/(\alpha - 1)} = \frac{\alpha - 1}{\alpha}(1 - t)^{-1/\alpha}.$$

Posons $q = (\alpha - 1)/\alpha > 0$. On a $q - 1 = -1/\alpha$, donc la fonction $t \mapsto 1 - (1 - t)^q$ possède exactement cette dérivée et vaut 0 en 0.

Comme $\mathcal{C}(0) = 0$, l'égalité des dérivées et des valeurs initiales donne $\mathcal{C}(t) = 1 - (1 - t)^q$ pour $t < 1$. La continuité donne ensuite la valeur en 1.

La première méthode utilise directement la définition de la courbe ; la seconde réemploie sa formule de dérivée générale.

1.c)

Les 30 % les plus riches possèdent 60 % de la masse des revenus. Les 70 % restants en possèdent donc 40 %. Par la définition de la courbe de concentration,

$$\mathcal{C}(0,7) = 0,4.$$

La formule précédente donne

$$1 - 0,3^{(\alpha-1)/\alpha} = 0,4,$$

soit $0,3^{(\alpha-1)/\alpha} = 0,6$. Les nombres concernés sont strictement positifs, donc on peut prendre le logarithme :

$$\frac{\alpha - 1}{\alpha} = \frac{\ln 0,6}{\ln 0,3}.$$

Comme $0,3 < 0,6 < 1$, le quotient des deux logarithmes appartient à $]0, 1[$. La valeur de α est donc bien strictement supérieure à 1, et elle est unique :

$$\alpha = \frac{\ln 0,3}{\ln 0,3 - \ln 0,6} = \frac{\ln 0,3}{\ln 0,5} \simeq 1,737.$$

Dans les questions suivantes, on revient à un paramètre quelconque $\alpha > 1$, conformément au sujet.

1.d)

Posons $q = (\alpha - 1)/\alpha \in]0, 1[$. La fonction de concentration est continue sur $[0, 1]$, donc l'intégrale définissant l'indice existe.

On a

$$I(X) = 2 \int_0^1 (t - 1 + (1 - t)^q) dt.$$

Une primitive de $(1 - t)^q$ est $-(1 - t)^{q+1}/(q + 1)$, et $q + 1 > 0$. Ainsi

$$\int_0^1 t dt = \frac{1}{2}, \quad \int_0^1 1 dt = 1, \quad \int_0^1 (1 - t)^q dt = \frac{1}{q + 1}.$$

Il vient

$$I(X) = -1 + \frac{2}{q+1}.$$

Comme $q+1 = (2\alpha-1)/\alpha$,

$$I(X) = \frac{1}{2\alpha-1}.$$

Pour $\alpha > 1$, cet indice appartient bien à $]0, 1[$. Il dépend du paramètre de forme α , mais pas du seuil x_0 .

2.a)

Posons $c = 1 - \lambda$. Comme $\lambda \in]0, 1[$, $c > 0$.

Méthode 1 — Stabilité de la loi de Pareto par changement d'échelle

D'après I.A.3, $Y = cX$ suit $VP(\alpha, cx_0)$. Le paramètre de forme est inchangé et $\alpha > 1$, donc la formule de 1.d) s'applique :

$$I(Y) = \frac{1}{2\alpha-1} = I(X).$$

Méthode 2 — Invariance directe de la courbe de concentration

Pour $y \geq cx_0$, la fonction de répartition de Y est $F_Y(y) = F_X(y/c)$ et son espérance vaut $cm > 0$.

La part des revenus inférieurs à y vaut

$$Q_Y(y) = \frac{E(cX \mathbf{1}_{\{X \leq y/c\}})}{cm} = \frac{E(X \mathbf{1}_{\{X \leq y/c\}})}{m} = Q_X(y/c).$$

Les variables tronquées sont intégrables par domination par X , donc ces espérances sont définies. La réciproque satisfait $F_Y^{-1}(t) = cF_X^{-1}(t)$ pour $t < 1$. Ainsi

$$\mathcal{C}_Y(t) = Q_Y(F_Y^{-1}(t)) = Q_X(F_X^{-1}(t)) = \mathcal{C}_X(t).$$

Par continuité, cette égalité vaut aussi en $t = 1$. Les deux indices, définis par la même intégrale, sont donc égaux.

La première méthode utilise la loi particulière de Pareto ; la seconde explique pourquoi une multiplication commune des revenus par une constante positive préserve les parts relatives.

2.b)

L'impôt proportionnel multiplie tous les revenus par le même facteur positif. Il ne change ni la courbe de concentration ni l'indice de Gini :

$$I(Y) = I(X).$$

L'inégalité des revenus mesurée par cet indice reste donc inchangée.

3.a)

Question adaptée au programme actuel — On précise $h > 0$. Si $h = 0$, tous les revenus disponibles seraient nuls et la masse totale des revenus serait nulle ; la courbe et l'indice définis dans le problème ne seraient pas applicables.

D'après I.A.6, $\sqrt{X} \sim VP(2\alpha, \sqrt{x_0})$. Puis, par I.A.3 et $h > 0$,

$$T = h\sqrt{X} \sim VP(2\alpha, h\sqrt{x_0}).$$

Le seuil est strictement positif et le paramètre de forme 2α est supérieur à 1, donc la formule de Gini s'applique :

$$I(T) = \frac{1}{4\alpha - 1}.$$

L'espérance nécessaire à la courbe existe ; elle vaut également $2\alpha h \sqrt{x_0} / (2\alpha - 1)$ d'après I.A.2.a).

3.b)

Pour $\alpha > 1$, les deux dénominateurs $2\alpha - 1$ et $4\alpha - 1$ sont strictement positifs, et $4\alpha - 1 > 2\alpha - 1$. Ainsi

$$I(T) = \frac{1}{4\alpha - 1} < \frac{1}{2\alpha - 1} = I(X).$$

La transformation progressive diminue donc l'inégalité mesurée par l'indice de Gini. Le facteur $h > 0$ ne joue que sur l'échelle des revenus et ne change pas cet indice.

4.a)

La variable $Z = X - a$ est une translation de X avec $\mu = -a$. D'après I.A.4.a),

$$Z \sim VP(\alpha, x_0 - a, a).$$

Le seuil $x_0 - a$ est strictement positif puisque $0 < a < x_0$, et la somme du seuil et du troisième paramètre vaut $x_0 > 0$.

La variable Z possède une densité continue strictement positive sur son intervalle de support et une espérance

$$E(Z) = m - a > 0,$$

car $m > x_0 > a$. Sa courbe de concentration est donc bien définie dans le cadre de la partie II.A.

4.b)

Méthode 1 — Transformation des revenus cumulés

Pour $z \geq x_0 - a$, $F_Z(z) = F_X(z + a)$, donc, pour $t \in [0, 1[$,

$$F_Z^{-1}(t) = F_X^{-1}(t) - a.$$

La densité de Z en z vaut $f_X(z + a)$. Dans l'intégrale des revenus cumulés, le changement de variable affine $u = z + a$ donne

$$Q_Z(z) = \frac{1}{m - a} \int_{x_0}^{z+a} (u - a) f_X(u) \, du.$$

Par linéarité,

$$Q_Z(z) = \frac{mQ_X(z + a) - aF_X(z + a)}{m - a}.$$

On compose avec F_Z^{-1} et on utilise $F_X(F_X^{-1}(t)) = t$:

$$\mathcal{C}_Z(t) = \frac{m\mathcal{C}(t) - at}{m - a}.$$

Cette égalité se prolonge en $t = 1$ par continuité. En la soustrayant de t ,

$$t - \mathcal{C}_Z(t) = \frac{(m - a)t - m\mathcal{C}(t) + at}{m - a} = \frac{m}{m - a}(t - \mathcal{C}(t)).$$

Méthode 2 — Dérivées des courbes de concentration

Les deux variables satisfont les hypothèses de II.A et $F_Z^{-1}(t) = F_X^{-1}(t) - a$. La formule de dérivée donne, pour $t < 1$,

$$\mathcal{C}'_Z(t) = \frac{F_X^{-1}(t) - a}{m - a} = \frac{m}{m - a} \mathcal{C}'(t) - \frac{a}{m - a}.$$

Les deux fonctions $\mathcal{C}_Z(t)$ et $(m\mathcal{C}(t) - at)/(m - a)$ ont ainsi la même dérivée sur $]0, 1[$ et valent toutes deux 0 en 0. Par le résultat sur les fonctions de dérivée nulle et la continuité en 0, elles sont égales sur $[0, 1[$, puis en 1 par continuité.

On retrouve la même identité de différences. La première méthode suit la masse des revenus après prélèvement; la seconde utilise la représentation par les quantiles obtenue dans la partie générale.

4.c)

Les courbes sont continues sur $[0, 1]$, donc les intégrales définissant les indices existent. Le facteur $m/(m - a)$ est une constante finie, dont le dénominateur est strictement positif.

Par 4.b) et linéarité de l'intégrale,

$$I(Z) = 2 \int_0^1 (t - \mathcal{C}_Z(t)) dt = \frac{m}{m - a} I(X).$$

En remplaçant m et $I(X)$,

$$I(Z) = \frac{\alpha x_0}{\alpha x_0 - a(\alpha - 1)} \frac{1}{2\alpha - 1}.$$

Le dénominateur $\alpha x_0 - a(\alpha - 1)$ est strictement positif, puisque $a < x_0$ implique qu'il est supérieur à x_0 .

4.d)

Comme $a > 0$ et $m > a$, on a $m/(m - a) > 1$. De plus, $I(X) = 1/(2\alpha - 1) > 0$. Ainsi

$$I(Z) > I(X).$$

Un prélèvement constant augmente donc l'inégalité mesurée par l'indice de Gini : il représente une proportion plus importante des petits revenus que des grands revenus.

La même conclusion se lit sur les courbes. D'après 4.b),

$$\mathcal{C}_Z(t) - \mathcal{C}(t) = \frac{a}{m - a}(\mathcal{C}(t) - t) \leq 0.$$

La courbe de concentration après cet impôt est donc située sous celle des revenus initiaux.

Relevé des adaptations

– I.B.5, graphique 1 : la figure est retranscrite en SVG dans l'énoncé à partir de l'original, avec les valeurs mesurées. Les neuf marqueurs sont placés avec les valeurs exactes du tableau, contrôlées sur l'image ; la droite D est reproduite à partir de mesures calibrées sur les graduations.

– I.B.5.2.a), I.B.5.2.b) et I.B.5.2.c) : le graphique 2 annoncé est absent du PDF fourni. Les neuf points sont reconstruits à partir du tableau avec $C_{\text{gr}} = 46$, valeur approchée compatible avec la lecture du premier graphique. La droite Δ est définie explicitement par les points associés à $x = 40$ et $x = 500$. La lecture d'une droite manquante est remplacée par le calcul de sa pente à partir de ces coordonnées. Les objectifs de linéarisation et d'estimation sont conservés.

– II.A.2.a), II.A.2.c) et II.A.2.d) : rectification de la dérivabilité à l'extrémité 1. La formule de dérivée vaut sur $[0, 1[$, avec une dérivée à droite en 0 ; en 1, la courbe est continue et possède une tangente verticale, et $F^{-1}(1)$ n'existe pas comme nombre réel.

– II.B.3.a) et II.B.3.b) : précision $h > 0$, nécessaire pour une masse totale de revenus strictement positive et pour l'utilisation de la loi de Pareto après changement d'échelle.