MAT 3775
Analyse de la régression

Chapitre 1
Préliminaires

P. Boily (uOttawa)

Session d'hiver – 2023

Aperçu

1 – Préliminaires

L'analyse de régression n'est pas une discipline très compliquée ... à condition de bien maîtriser ses pré-requis. Dans ce cours, il sera utile de se familiariser avec un certain nombre de notions relatives :

- aux **variables aléatoires** ;
- au **calcul à plusieurs variables** ;
- à l'**algèbre linéaire** ;
- aux **formes quadratiques**, et
- à l'**optimisation**.

1.1 – Variables aléatoires

Une **épreuve aléatoire** est un **processus** pour lequel il est impossible de prédire le **résultat avec certitude**. L'**espace d'échantillonnage** $\mathcal{S}$ est l'ensemble des **résultats possibles** de l'épreuve aléatoire.

Une **variable aléatoire** Y associée est une **fonction** $Y : \mathcal{S} \to \mathbb{R}$. Si l'ensemble $Y(\mathcal{S}) = \{Y(s) \mid s \in \mathcal{S}\}$ est **dénombrable**, Y est une **variable aléatoire discrète** ; s'il ne l'est pas, Y est une **variable aléatoire continue**.

À chaque v.a. Y correspond une **fonction de probabilité** $f(Y)$, qui spécifie les **probabilités des valeurs prises par** Y.

Y_1 et Y_2 sont **indépendantes** lorsque leur **f.d.p. conjointe** $f(Y_1, Y_2)$ est le produit des **f.d.p. individuelles** $f(Y_1)f(Y_2)$.

1.1.1 – Espérance, variance, et covariance

L'**opérateur d'espérance** $\mathrm{E}\{\cdot\}$ est défini par

$$\mathrm{E}\{Y\} = \begin{cases} \displaystyle\sum_{Y(s)} Y(s) f(Y(s)), & \text{si } Y \text{ est discrète} \\ \displaystyle\int_{\mathbb{R}} Y f(Y)\, dy, & \text{si } Y \text{ est continue} \end{cases}$$

L'espérance $\mathrm{E}\{Y\}$ est la **valeur moyenne** que l'on s'attend à observer si l'expérience est répétée à maintes reprises.

L'espérance est parfois aussi appelée la **moyenne** de Y, notée $\overline{Y}$; c'est donc une mesure de la **tendance centrale** de Y.

L'**opérateur de variance** $\sigma^2\{\cdot\}$ est défini par

$$\sigma^2\{Y\} = \mathrm{E}\left\{(Y - \mathrm{E}\{Y\})^2\right\} = \mathrm{E}\{Y^2\} - \left(\mathrm{E}\{Y\}\right)^2.$$

Il est souvent désigné par Var(Y). C'est une mesure de la **dispersion** de Y (les grandes variances sont associées à de **fortes dispersions**, et *vice-versa*).

L'opérateur de **covariance** $\sigma\{\cdot,\cdot\}$ est défini par

$$\sigma\{Y,W\} = \mathrm{E}\{(Y - \mathrm{E}\{Y\})(W - \mathrm{E}\{W\})\} = \mathrm{E}\{YW\} - \mathrm{E}\{Y\}\mathrm{E}\{W\}.$$

Il est souvent désigné par Cov(Y,W). C'est une mesure de la **force de la relation linéaire** entre deux v.a. (les grandes magnitudes de covariance sont associées à la **linéarité**, mais "grand" est un concept relatif).

L'opérateur **écart-type** $\sigma\{\cdot\}$ est défini par

$$\sigma\{Y\} = \sqrt{\sigma^2\{Y\}}.$$

Il est toujours non négatif.

L'**opérateur de corrélation** $\rho\{\cdot,\cdot\}$ est défini par

$$\rho\{Y, W\} = \frac{\sigma\{Y, W\}}{\sigma\{Y\}\sigma\{W\}},$$

en supposant que $\sigma\{Y\}\sigma\{W\} \neq 0$.

Lorsque $\rho\{Y, W\} = 0$, on dit que les v.a. sont **non corrélées**.

Propriétés des opérateurs

Soient Y, Y_i, W des v.a., et $a, b, c, a_i, b_i, c_i \in \mathbb{R}$, $i = 1, \ldots, n$. Alors :

- $\mathrm{E}\{\cdot\}$ est **linéaire** sur l'espace de v.a. : $\mathrm{E}\{aY + b\} = a\mathrm{E}\{Y\} + b$ et

$$\mathrm{E}\left\{\sum_{i=1}^{n} a_i Y_i\right\} = \sum_{i=1}^{n} a_i \mathrm{E}\{Y_i\}$$

- $\sigma^2\{aY + b\} = a^2\sigma^2\{Y\}$ et

$$\sigma^2\left\{\sum_{i=1}^{n} a_i Y_i\right\} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma\{Y_i, Y_j\} = \sum_{i=1}^{n} a_i^2 \sigma^2\{Y_i\} + \sum_{i \neq j} a_i a_j \sigma\{Y_i, Y_j\}$$

- $\sigma\{Y,Y\} = \sigma^2\{Y\}$ et $\sigma\{Y,W\} = \sigma\{W,Y\}$
- $\sigma\{a_1Y + b_1, a_2W + b_2\} = a_1a_2\sigma\{Y,W\}$
- $\{Y_i\}$ **sans corrélation** $\implies$

$$\sigma\left\{\sum_{i=1}^n a_iY_i, \sum_{i=1}^n c_iY_i\right\} = \sum_{i=1}^n a_ic_i\sigma^2\{Y_i\}$$

- $\sigma\{Y,W\} < 0 \iff$ les observations de Y au-dessus de $\overline{Y}$ ont tendance à accompagner les observations de W **en dessous** de $\overline{W}$, et *vice-versa*
- $\sigma\{Y,W\} > 0 \iff$ les observations de Y au-dessus de $\overline{Y}$ ont tendance à accompagner les observations de W **au-dessus** de $\overline{W}$, et *vice-versa*

- $\sigma\{Y,W\}=0 \implies Y$ et W sont **sans corrélation**
- Y,W **indépendantes** $\implies \rho\{Y,W\}=0$ (sans corrélation)
- $\rho\{Y,W\}=0 \not\Longrightarrow Y,W$ **indépendantes**, cependant
- $|\rho\{Y,W\}|\leq 1$ (conséquence de l'inégalité de Cauchy-Schwartz)
- $|\rho\{Y,W\}|=1 \iff Y=aW+b$ pour une paire $(a,b)\in\mathbb{R}^2$,

Vecteurs aléatoires

Si $Y_1, \ldots, Y_n$ sont des variables aléatoires, alors

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

est un **vecteur aléatoire**. L'**espérance** de $\mathbf{Y}$ est

$$\mathrm{E}\left\{\mathbf{Y}\right\} = \begin{pmatrix} \mathrm{E}\left\{Y_1\right\} \\ \vdots \\ \mathrm{E}\left\{Y_n\right\} \end{pmatrix}.$$

Les composantes de $\mathbf{Y}$ n'ont pas nécessairement tous des distributions **identiques**.

La **matrice de variance-covariance** de $\mathbf{Y}$ est la matrice symétrique

$$\sigma^2\{\mathbf{Y}\} = (g_{i,j}), \quad \text{où } g_{i,j} = \begin{cases} \sigma^2\{Y_i\} & i = j \\ \sigma\{Y_i, Y_j\} & i \neq j \end{cases}$$

ou encore

$$\sigma^2\{\mathbf{Y}\} = \begin{pmatrix} \sigma^2\{Y_1\} & \cdots & \sigma\{Y_1, Y_n\} \\ \vdots & \ddots & \vdots \\ \sigma\{Y_1, Y_n\} & \cdots & \sigma^2\{Y_n\} \end{pmatrix}$$

Si les composantes de $\mathbf{Y}$ sont **indépendantes** et ont toutes la **même variance** σ^2, alors

$$\sigma^2\{\mathbf{Y}\} = \sigma^2\mathbf{I}_n.$$

En pratique, nous travaillons généralement avec des **échantillons** de v.a. Si l 'on observe $\{(X_i, Y_i)\}_{i=1}^n$ à partir de la distribution conjointe de (X, Y) :

- $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ et $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$, les **moyennes d'échantillon**, sont des estimateurs sans biais de $\mathrm{E}\{X\}$ et $\mathrm{E}\{Y\}$, respectivement ;
- $s_X^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ et $s_Y^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$, les **variances d'échantillon**, sont des estimateurs sans biais de $\sigma^2\{X\}$ et $\sigma^2\{Y\}$;
- $s_{XY} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})$, la **covariance d'échantillon**, est un estimateur sans biais de $\sigma\{X, Y\}$.

1.1.2 – Distributions importantes

La **fonction de répartition (cumulative)** (f.r.c.) de toute variable aléatoire continue Y est définie par

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} f_Y(t)\,dt$$

que l'on considère comme fonction d'une variable réelle y. Alternativement, nous pouvons décrire la **loi** (distribution) de Y *via* la relation suivante entre $f_Y(y)$ et $F_Y(y)$:

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

Fonction de probabilité

La **fonction de densité de probabilité** (f.d.p.) d'une variable aléatoire continue Y est une fonction **intégrable** $f_Y : Y(\mathcal{S}) \to \mathbb{R}$ telle que :

- $f_Y(y) > 0$ pour tout $y \in Y(\mathcal{S})$ et $\lim_{y\to\pm\infty} f_Y(y) = 0$;
- $\int_{\mathcal{S}} f_Y(y)\, dy = 1$;
- pour toute paire $(a, b) \in \mathbb{R}^2$,

$$\begin{aligned} P(a < Y < b) &= P(a \leq Y < b) = P(a < Y \leq b) = P(a \leq Y \leq b) \\ &= F_Y(b) - F_Y(a) = \int_a^b f(y)\, dy. \end{aligned}$$

Loi normale : la f.r.c. de la v.a. $Y \sim \mathcal{N}(\mu, \sigma^2)$ est

$$F_Y(y) = P(Y \leq y) = \Phi(y),$$

avec

$$f_Y(y) = \Phi'(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2 \right).$$

Loi χ^2 : la f.d.p. de la v.a. $Y \sim \chi^2(\nu)$ est

$$f_Y(y;\nu) = \begin{cases} \dfrac{y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}, & y > 0; \\ 0, & \text{autrement} \end{cases}$$

où $\Gamma(\cdot)$ représente la **fonction Gamma**.

Si $U_i \sim \chi^2(\nu_i)$, $i = 1, 2$, et U_1, U_2 sont indépendantes, alors

$$U = U_1 + U_2 \sim \chi^2(\nu_1) + \chi^2(\nu_2) = \chi^2(\nu_1 + \nu_2).$$

Il existe un lien important entre la loi normale centrée réduite $\mathcal{N}(0, 1)$ et la loi $\chi^2(1)$: si $Z \sim \mathcal{N}(0, 1)$, alors $Z^2 \sim \chi^2(1)$.

Loi de Student : si $Z \sim \mathcal{N}(0, 1)$ et $U \sim \chi^2(\nu)$, Z, U indépendantes :

$$t = \frac{Z}{\sqrt{U/\nu}} \sim t(\nu),$$

suit une **loi T de Student avec ν degrés de liberté**.

Loi de Fisher : si $U_i \sim \chi^2(\nu_i)$, $i = 1, 2$ et U_1, U_2 indépendantes :

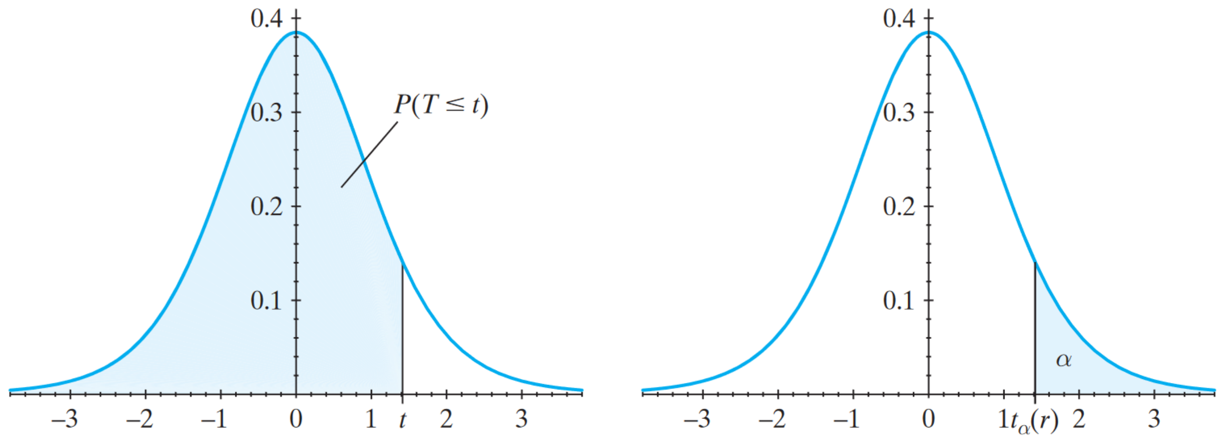
$$F = \frac{U_1/\nu_1}{U_2/\nu_2} \sim F(\nu_1, \nu_2),$$

suit une loi de **Fisher avec ν_1 et ν_2 degrés de liberté**.

Je vous encourage à savoir lire les tableaux de f.r.c. et les fonctions `R` correspondantes :

- `qt()`, `dt()`, `pt()`, `rt()`, et
- `qf()`, `df()`, `pf()`, `rf()`.

Table VI The *t* Distribution



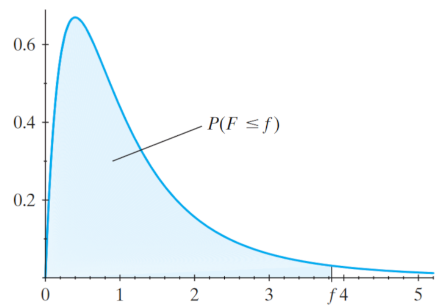
$$P(T \le t) = \int_{-\infty}^{t} \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma(r/2)(1+w^2/r)^{(r+1)/2}}\,dw$$

$$P(T \le -t) = 1 - P(T \le t)$$

	$P(T \le t)$						
	0.60	0.75	0.90	0.95	0.975	0.99	0.995
r	$t_{0.40}(r)$	$t_{0.25}(r)$	$t_{0.10}(r)$	$t_{0.05}(r)$	$t_{0.025}(r)$	$t_{0.01}(r)$	$t_{0.005}(r)$
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841

Table VII The F Distribution

$$P(F \le f) = \int_0^f \frac{\Gamma[(r_1+r_2)/2](r_1/r_2)^{r_1/2} w^{r_1/2-1}}{\Gamma(r_1/2)\Gamma(r_2/2)(1+r_1 w/r_2)^{(r_1+r_2)/2}}\, dw$$



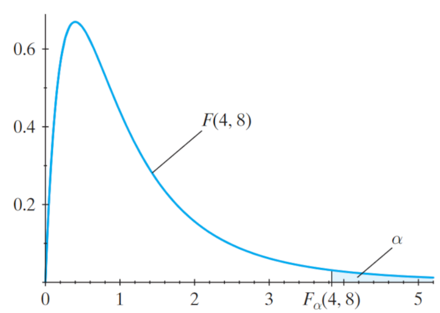


Table VII *continued*

$$P(F \le f) = \int_0^f \frac{\Gamma[(r_1+r_2)/2](r_1/r_2)^{r_1/2} w^{r_1/2-1}}{\Gamma(r_1/2)\Gamma(r_2/2)(1+r_1 w/r_2)^{(r_1+r_2)/2}}\, dw$$

		Den. d.f.	Numerator Degrees of Freedom, r_1									
α	$P(F \le f)$	r_2	1	2	3	4	5	6	7	8	9	10
0.05	0.95	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
0.025	0.975		647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63
0.01	0.99		4052	4999.5	5403	5625	5764	5859	5928	5981	6022	6056
0.05	0.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
0.025	0.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
0.01	0.99		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
0.05	0.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
0.025	0.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
0.01	0.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
0.05	0.95	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
0.025	0.975		12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84
0.01	0.99		21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
0.05	0.95	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
0.025	0.975		10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
0.01	0.99		16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
0.05	0.95	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
0.025	0.975		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46
0.01	0.99		13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
0.05	0.95	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
0.025	0.975		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76
0.01	0.99		12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62

Théorème central limite

Théorème : soient $X_1, \ldots, X_n$ des v.a. normales indépendantes de moyennes $\mu_1, \ldots, \mu_n$ et d'écart-types $\sigma_1, \ldots, \sigma_n$. Alors

$$X_1 + \cdots + X_n \sim \mathcal{N}(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2).$$

Si $\mu_i \equiv \mu$ et $\sigma_i^2 \equiv \sigma$ pour $i = 1, \ldots, n$, alors $X_1 + \cdots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$.

Théorème : soient $X_1, \ldots, X_n$ des v.a. normales indépendantes de moyenne μ et d'écart-type σ. Soit $\overline{X}$ la moyenne d'échantillon. Alors

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Théorème : soient $X_1, \ldots, X_n$ des v.a. indépendantes de moyenne μ et d'écart-type σ. Soit $\overline{X}$ la moyenne d'échantillon. Alors

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \to Z \sim \mathcal{N}(0,1), \quad \text{lorsque } n \to \infty.$$

soient $X_1, \ldots, X_n$ des v.a. normales indépendantes de moyenne μ et d'écart-type σ **inconnu**. Soit $\overline{X}$ et S^2 la moyenne et la variance d'échantillon, respectivement. Alors

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

suit une **loi T de Student avec $\nu = n - 1$ degrés de liberté**.

1.2 – Calcul multivarié

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction **différentiable**. Si $\mathbf{Y} = (Y_1, \ldots, Y_n)$, la **dérivée** de f par rapport à $\mathbf{Y}$ est

$$\nabla_{\mathbf{Y}} f(\mathbf{Y}) = \begin{pmatrix} \dfrac{\partial f(\mathbf{Y})}{\partial Y_1} \\ \vdots \\ \dfrac{\partial f(\mathbf{Y})}{\partial Y_n} \end{pmatrix}.$$

Le **gradient** ∇ est un **opérateur linéaire** :

$$\nabla_{\mathbf{Y}}(af + bg)(\mathbf{Y}) = a\nabla_{\mathbf{Y}} f(\mathbf{Y}) + b\nabla_{\mathbf{Y}} g(\mathbf{Y}).$$

If $f(\mathbf{Y}) \equiv a$, alors $\nabla_{\mathbf{Y}} f(\mathbf{Y}) = \mathbf{0}$. Si $f(\mathbf{Y}) = \mathbf{Y}^\top \mathbf{v}$, alors $\nabla_{\mathbf{Y}} f(\mathbf{Y}) = \mathbf{v}$.

1.3 – Algèbre matricielle

Soient $A \in M_{m,n}(\mathbb{R})$ et $\mathbf{Y}$ un vecteur aléatoire. Si $\mathbf{W} = A\mathbf{Y}$, alors

$$\mathrm{E}\{\mathbf{W}\} = A\mathrm{E}\{\mathbf{Y}\} \quad \text{et} \quad \sigma^2\{\mathbf{W}\} = A\sigma^2\{\mathbf{Y}\}A^\top.$$

De plus, si $\mathbf{Y} \sim \mathcal{N}\left(\mathrm{E}\{\mathbf{Y}\}, \sigma^2\{\mathbf{Y}\}\right)$, alors

$$\mathbf{W} \sim N\left(\mathrm{E}\{\mathbf{W}\}, \sigma^2\{\mathbf{W}\}\right) = \mathcal{N}\left(A\mathrm{E}\{\mathbf{Y}\}, A\sigma^2\{\mathbf{Y}\}A^\top\right).$$

Si $A \in M_{n,n}(\mathbb{R})$, la **trace** de A est $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}$.

La trace est un **opérateur linéaire** : $\mathrm{tr}(kA + B) = k\,\mathrm{tr}\,A + \mathrm{tr}\,B$; on a aussi $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ (lorsque les matrices sont **compatibles**).

La **transposition** d'une matrice A, notée $A^\top$, est obtenue en interchangeant ses **lignes** et ses **colonnes**, ou simplement en **réfléchissant** la matrice le long de sa **diagonale principale**.

Propriétés: si $A \in M_{m,n}(\mathbb{R})$ et $k \in \mathbb{R}$, alors

- $(A^\top)^\top = A$
- $k^\top = k$
- $(kA + B)^\top = kA^\top + B^\top$
- $(AB)^\top = B^\top A^\top$

1.4 – Formes quadratiques

Une **forme quadratique symétrique** en $Y_1, \ldots, Y_n$ est une expression de la forme

$$Q_A(\mathbf{Y}) = \mathbf{Y}^\top A\mathbf{Y} = \sum_{i,j=1}^{n} a_{i,j}Y_iY_j,$$

où A est une **matrice symétrique** $n \times n$ ($A^\top = A$).

Un certain nombre de quantités importantes apparaissant en analyse de régression peuvent être exprimées sous de telles formes.

Les **degrés de liberté** d'une forme quadratique symétrique $Q_A(\mathbf{Y})$ peuvent être obtenus en calculant le **rang** de la matrice associée A.

Par exemple, la matrice symétrique associée à la forme quadratique symétrique $Q_A(\mathbf{Y}) = 4Y_1^2 + 7Y_1Y_2 + 2Y_2^2$ est

$$A = \begin{pmatrix} 4 & 7/2 \\ 7/2 & 2 \end{pmatrix}; \quad Q_A \text{ a } 2 \text{ degrés de liberté.}$$

Théorème: soient $Q_1, \ldots Q_K$ des formes quadratiques symétriques en $\mathbf{Y}$ avec matrices symétriques associées $A_1, \ldots, A_K$, respectivement. Si $a_i \in \mathbb{R}$ pour tout $i = 1, \ldots, K$, alors

$$Q = a_1 Q_1 + \cdots + a_K Q_K$$

est une forme quadratique symétrique en $\mathbf{Y}$ avec matrice symétrique associée

$$A = a_1 A_1 + \cdots + a_K A_K.$$

Pour une matrice B générale de dimension $n \times n$, nous avons

$$\nabla_{\mathbf{Y}} \left(\mathbf{Y}^{\top} B \mathbf{Y} \right) = (B^{\top} + B)\mathbf{Y}.$$

Ainsi le gradient d'une forme quadratique symétrique $Q_A(\mathbf{Y})$ est

$$\nabla_{\mathbf{Y}} Q_A(\mathbf{Y}) = 2A\mathbf{Y}.$$

On peut montrer que **chaque** expression de la forme $\mathbf{Y}^{\top} B \mathbf{Y}$ peut être associée à une matrice symétrique A. Nous supposons donc que chaque forme de ce type est symétrique.

Le rôle joué par les formes quadratiques dans le calcul multi-variable est analogue au rôle joué par $f(x) = ax^2$ dans le calcul différentiel ordinaire.

Les **valeurs propres** d'une matrice A de taille $n \times n$ sont les racines du **polynôme caractéristique** $p_A(\lambda)$ de A : $p_A(\lambda) = \det(A - \lambda \mathbf{I}_n) = 0$.

Il existe n telles racines (complexes), pas nécessairement distinctes.

Si λ est une valeur propre de A, alors $\exists \mathbf{v} \neq \mathbf{0}$ tel que $A\mathbf{v} = \lambda \mathbf{v}$. Si de plus A est réelle et symétrique, toutes ses valeurs propres sont **réelles**.

Soit $Q_A(\mathbf{Y})$ une forme quadratique avec $\text{eig}(A) = \{\lambda_1, \ldots, \lambda_n\} \subseteq \mathbb{R}$:

- si $\lambda_i > 0$ pour tout i, $Q_A(\mathbf{Y})$ et A sont dîtes **définies positives** ;
- si $\lambda_i < 0$ pour tout i, $Q_A(\mathbf{Y})$ et A sont dîtes **définies négatives** ;
- si $\lambda_i \lambda_j < 0$ pour une paire i, j quelconque, $Q_A(\mathbf{Y})$ et A sont **indéfinies**.

1.4.1 – Théorème de Cochran

Soient $\mathbf{Y} = (Y_1, \ldots, Y_n) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

Supposons que
$$\mathbf{Y}^\top \mathbf{Y} = Q_1(\mathbf{Y}) + \cdots + Q_K(\mathbf{Y}),$$
où les Q_k sont des formes quadratiques (semi-)définies positives avec $r_k (= \mathrm{rang}(A_k))$ degrés de liberté, pour $k = 1, \ldots, K$.

Si $r_1 + \cdots + r_K = n$, alors $Q_1(\mathbf{Y}), \ldots, Q_K(\mathbf{Y})$ sont **indépendantes** et

$$\frac{Q_k(\mathbf{Y})}{\sigma^2} \sim \chi^2(r_k), \quad k = 1, \ldots, K.$$

En particulier, si $K = 2$ et $r_1 = r$, alors $Q_2(\mathbf{Y})/\sigma^2 \sim \chi^2(n-r)$.

1.4.2 – Formes quadratiques importantes

Pour tout entier positif n, nous définissons deux **matrices spéciales** :

$$\mathbf{J}_n = \mathbf{J} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \quad \text{et} \quad \mathbf{1}_{n\times 1} = \mathbf{1}_n = \mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

On note que $\mathbf{1}_n^\top \mathbf{1}_n = n$ et $\mathbf{1}_n \mathbf{1}_n^\top = \mathbf{J}_n$. Soit $\mathbf{Y} = (Y_1, \ldots, Y_n) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ un vecteur aléatoire. Quelles sont les matrices symétriques associées à

$$Q_A(\mathbf{Y}) = \sum_{i=1}^n Y_i^2, \quad Q_B(\mathbf{Y}) = n\overline{Y}^2, \quad \text{et} \quad Q_C(\mathbf{Y}) = \sum_{i=1}^n (Y_i - \overline{Y})^2?$$

Ré-écrivons les formes quadratiques en $\mathbf{Y}$ pour obtenir (page suivante) :

$$Q_A(\mathbf{Y}) = \mathbf{Y}^\top\mathbf{Y} = \mathbf{Y}^\top\mathbf{I}_n\mathbf{Y} \implies A = \mathbf{I}_n;$$

$$Q_B(\mathbf{Y}) = n\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)^2 = \frac{1}{n}\sum_{i,j=1}^n Y_iY_j = \frac{1}{n}\mathbf{Y}^\top\mathbf{1}_n\mathbf{1}_n^\top\mathbf{Y} \implies B = \frac{1}{n}\mathbf{J}_n;$$

$$Q_C(\mathbf{Y}) = \sum_{i=1}^n Y_i^2 - n\overline{Y}^2 = \mathbf{Y}^\top\mathbf{I}_n\mathbf{Y} - \frac{1}{n}\mathbf{Y}^\top\mathbf{J}_n\mathbf{Y} \implies C = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n.$$

Puisque $\mathrm{rang}(A) = n$, $\mathrm{rang}(B) = 1$, et $\mathrm{rang}(C) = n-1$, le théorème de Cochran implique que $Q_A(\mathbf{Y})$, $Q_B(\mathbf{Y})$, et $Q_C(\mathbf{Y})$ sont des v.a. **indépendantes**, et que

$$\frac{Q_A(\mathbf{Y})}{\sigma^2} = \frac{\mathbf{Y}^\top\mathbf{Y}}{\sigma^2} \sim \chi^2(n), \; \frac{Q_B(\mathbf{Y})}{\sigma^2} = \frac{n\overline{Y}^2}{\sigma^2} \sim \chi^2(1), \; \frac{Q_C(\mathbf{Y})}{\sigma^2} = \frac{\mathrm{SST}}{\sigma^2} \sim \chi^2(n-1).$$

1.5 – Optimisation

Soient A une matrice symétrique $n \times n$, $\mathbf{v} \in \mathbb{R}^n$, $c \in \mathbb{R}$, et

$$f(\mathbf{Y}) = \frac{1}{2}\mathbf{Y}^\top A\mathbf{Y} - \mathbf{Y}^\top \mathbf{v} + c.$$

Notons que f est **dérivable**. Les **points critiques** de f satisfont à

$$\nabla_{\mathbf{Y}} f(\mathbf{Y}) = A\mathbf{Y} - \mathbf{v} = \mathbf{0} \implies A\mathbf{Y} = \mathbf{v}.$$

Si A est **inversible** ($\det(A) \neq 0$), le point critique $\mathbf{Y}^* = A^{-1}\mathbf{v}$ est **unique**.

Si A est **singulière** ($\det(A) = 0$), il y a soit **pas** de point critique (si $\mathbf{v} \notin \mathbf{image}(A)$) ou une **infinité** de points critiques (si $\mathbf{v} \in \mathbf{image}(A)$).

Lorsque A est **inversible**:

- si A est **définie positive**, alors f atteint son **minimum global** en $\mathbf{Y}^* = A^{-1}\mathbf{v}$;
- si A est **définie négative**, alors f atteint son **maximum global** en $\mathbf{Y}^* = A^{-1}\mathbf{v}$;
- si A est **indéfinie** (c'est-à-dire que A possède des valeurs propres positives **et** des valeurs propres négatives), alors $\mathbf{Y}^* = A^{-1}\mathbf{v}$ est un **col** de f.

Si les valeurs propres peuvent être **nulle**, nous remplaçons "défini" par "semi-défini".