Eléments du Calcul des Probabilités
Chapitre 4: Ensembles de variables aléatoires et conditionnement

Louis Wehenkel

Département d'Electricité, Electronique et Informatique - Université de Liège
B24/II.93 - L.Wehenkel@ulg.ac.be

MATH0062-1 : 2BacIng, 2BacInf - 19/4/2017

Find slides: http://montefiore.ulg.ac.be/∼lwh/Probas/

Remarques concernant le syllabus...

- Le chapitre 3 a été légèrement modifié : quelques coquilles ont été corrigées et quelques éléments ont été ajoutés (voir erratum, à la fin du syllabus pour une indication en ce qui concerne les changements opérés).
- Toutes remarques, questions, sur le syllabus sont toujours souhaitées.
- Rappel: chapitre 5 toujours incomplet à ce stade.
- Quelques remarques concernant les projets...

4.1 Variables aléatoires discrètes et conditionnement

- Définitions générales
  - Lois conjointes, marginales, et conditionnelles
- Variables à valeurs réelles
  - Espérance et variance conditionnelles
- Espace de variables aléatoires réelles définies sur un Ω fini
  - Première analyse de la géométrie des v.a. à v.r.

Couples de v.a. : définitions générales

Hypothèses

- On considère un espace de probabilités $(\Omega, \mathcal{E}, P)$
- On considère deux variables aléatoires discrètes $\mathcal{X}$ et $\mathcal{Y}$
- Pour simplifier, nous supposons que $\Omega_\mathcal{X}$ et $\Omega_\mathcal{Y}$ sont finis, $\Omega_\mathcal{X} = \{x_1, \ldots, x_k\}$ et $\Omega_\mathcal{Y} = \{y_1, \ldots, y_l\}$ et que ces ensembles sont munis de leur $\sigma-$algèbre maximale $\mathcal{E}_\mathcal{X} = 2^{\Omega_\mathcal{X}}$ et $\mathcal{E}_\mathcal{Y} = 2^{\Omega_\mathcal{Y}}$.
- Les variables $\mathcal{X}$ et $\mathcal{Y}$ induisent les densités (et mesures) de probabilité $P_\mathcal{X}$ sur $(\Omega_\mathcal{X}, \mathcal{E}_\mathcal{X})$ et $P_\mathcal{Y}$ sur $(\Omega_\mathcal{Y}, \mathcal{E}_\mathcal{Y})$ (cf chap. 3).
- Observer la valeur du couple de variables $\mathcal{X}$ et $\mathcal{Y}$ revient à observer la valeur de la v.a. $\mathcal{Z}$, définie par $\mathcal{Z}(\omega) \stackrel{\triangle}{=} (\mathcal{X}(\omega), \mathcal{Y}(\omega)), \forall \omega \in \Omega$, avec $\Omega_\mathcal{Z} = \Omega_\mathcal{X} \times \Omega_\mathcal{Y}$, et $\mathcal{E}_\mathcal{Z} = 2^{\Omega_\mathcal{Z}} = \mathcal{E}_\mathcal{X} \otimes \mathcal{E}_\mathcal{Y}$, et de densité (et mesure) $P_\mathcal{Z}$.

NB: dans la suite nous utilisons la notation $P_\mathcal{Z}$ pour désigner la densité de la variable $\mathcal{Z}$, c'est-à-dire la fonction définie sur $\Omega_\mathcal{Z}$ qui associe à une valeur $z \in \Omega_\mathcal{Z}$ la probabilité de l'événement $\{\omega : \mathcal{Z}(\omega) = z\}$. Cette densité permet de calculer la mesure de probabilité pour tout sous-ensemble Z de $\Omega_\mathcal{Z}$, par sommation sur Z.

Couples de v.a. : loi conjointe

La loi (densité) de probabilité conjointe $P_{\mathcal{X},\mathcal{Y}}$ du couple $(\mathcal{X},\mathcal{Y})$ est déterminée complètement par la connaissance des kl nombres

$$P_{\mathcal{X},\mathcal{Y}}(x_i, y_j) \triangleq P([\mathcal{X} = x_i] \wedge [\mathcal{Y} = y_j]), \forall i = 1, \ldots, k, \forall j = 1, \ldots, l. \quad (1)$$

$P_{\mathcal{X},\mathcal{Y}}(x_i, y_j)$ désigne donc la probabilité de l'événement $\{\omega \in \Omega : \mathcal{X}(\omega) = x_i \text{ et } \mathcal{Y}(\omega) = y_j\}$.

En effet, une fois qu'on connait la densité $P_{\mathcal{X},\mathcal{Y}}(x_i, y_j)$ pour toutes les valeurs de ses arguments, on peut en déduire la probabilité de tout événement qui peut se décrire par une affirmation logique faisant intervenir les valeurs des deux variables aléatoires.

On a bien entendu que $\sum_{i=1}^{k} \sum_{j=1}^{l} P_{\mathcal{X},\mathcal{Y}}(x_i, y_j) = 1$.

Couples de v.a. : déduction des lois marginales

La loi (densité) de $\mathcal{X}$ est obtenue à partir de la loi conjointe par l'opération de **marginalisation** :

$$P_{\mathcal{X}}(x_i) \triangleq P(\mathcal{X} = x_i) = \sum_{j=1}^{l} P_{\mathcal{X},\mathcal{Y}}(x_i, y_j), \forall i = 1, \ldots k. \quad (2)$$

De même, la loi (densité) de $\mathcal{Y}$ est obtenue par

$$P_{\mathcal{Y}}(y_j) \triangleq P(\mathcal{Y} = y_j) = \sum_{i=1}^{k} P_{\mathcal{X},\mathcal{Y}}(x_i, y_j), \forall j = 1, \ldots l. \quad (3)$$

On désigne souvent ces lois par le terme de **lois (densités) marginales**.

Ces formules sont une conséquence du théorème des probabilités totales.

Couples de v.a. : table de contingences

On représente souvent un couple de v.a. discrètes à l'aide d'une *table de contingences*, comme illustré à la Figure.

	y_1	$\cdots$	y_j	$\cdots$	y_l	$\sum$
x_1						
$\vdots$			$\vdots$			
x_i		$\cdots$	$P_{\mathcal{X},\mathcal{Y}}(x_i, y_j)$	$\cdots$		$P_{\mathcal{X}}(x_i)$
$\vdots$			$\vdots$			
x_k						
$\sum$			$P_{\mathcal{Y}}(y_j)$			1

Figure : Table de contingences (et densité conjointe)

Couples de v.a. : lois conditionnelles

La loi (densité) conditionnelle de $\mathcal{X}$ connaissant $\mathcal{Y}$ est définie par

$$P_{\mathcal{X}|\mathcal{Y}}(x_i|y_j) \triangleq P(\mathcal{X}=x_i|\mathcal{Y}=y_j) = \frac{P_{\mathcal{X},\mathcal{Y}}(x_i,y_j)}{P_{\mathcal{Y}}(y_j)}, \forall i=1,\ldots k, \forall j=1,\ldots l. \quad (4)$$

La loi (densité) conditionnelle de $\mathcal{Y}$ connaissant $\mathcal{X}$ est définie par

$$P_{\mathcal{Y}|\mathcal{X}}(y_j|x_i) \triangleq P(\mathcal{Y}=y_j|\mathcal{X}=x_i) = \frac{P_{\mathcal{X},\mathcal{Y}}(x_i,y_j)}{P_{\mathcal{X}}(x_i)}, \forall i=1,\ldots k, \forall j=1,\ldots l. \quad (5)$$

Ces lois conditionnelles ne sont ainsi définies que pour les valeurs de y_j (resp. de x_i) de probabilité marginale non-nulle. Cependant, la valeur précise qu'on leur attribue lorsque $P_{\mathcal{Y}}(y_j)=0$, (resp. $P_{\mathcal{X}}(x_i)=0$) auquel cas on a aussi $P_{\mathcal{X},\mathcal{Y}}(x_i,y_j)=0$, n'a pas réellement d'importance; on choisit alors arbitrairement de fixer leur valeur à 0.

Couples de v.a. : factorisation de la loi conjointe

On a, $\forall x \in \Omega_{\mathcal{X}}, \forall y \in \Omega_{\mathcal{Y}}$, $P_{\mathcal{X},\mathcal{Y}}(x,y) = P_{\mathcal{X}|\mathcal{Y}}(x|y)P_{\mathcal{Y}}(y) = P_{\mathcal{Y}|\mathcal{X}}(y|x)P_{\mathcal{X}}(x)$.

Nous écrivons cela de façon plus compacte par

Factorisation de la loi (densité) conjointe

$$P_{\mathcal{X},\mathcal{Y}} = P_{\mathcal{X}|\mathcal{Y}}P_{\mathcal{Y}} = P_{\mathcal{Y}|\mathcal{X}}P_{\mathcal{X}}.$$

Remarquons que lorsque $\mathcal{X} \perp \mathcal{Y}$, on a

$$\forall x \in \Omega_{\mathcal{X}}, \forall y \in \Omega_{\mathcal{Y}} : P_{\mathcal{X}|\mathcal{Y}}(x|y) = P_{\mathcal{X}}(x), \text{ et}$$

$$\forall x \in \Omega_{\mathcal{X}}, \forall y \in \Omega_{\mathcal{Y}} : P_{\mathcal{Y}|\mathcal{X}}(y|x) = P_{\mathcal{Y}}(y), \text{ et}$$

par conséquent $\forall x \in \Omega_{\mathcal{X}}, \forall y \in \Omega_{\mathcal{Y}} : P_{\mathcal{X},\mathcal{Y}}(x,y) = P_{\mathcal{X}}(x)P_{\mathcal{Y}}(y)$, ce que nous écrivons

$$P_{\mathcal{X},\mathcal{Y}} = P_{\mathcal{X}}P_{\mathcal{Y}}.$$

Exemple 2: lois de probabilité induites (a)

NB: la mesure P définie sur Ω est uniforme ($P(\omega) = 1/36, \forall \omega \in \Omega$).

- Pour rappel, on a
  - $P_{\mathcal{Y}_1}(y_1) = 1/6, \forall y_1 \in \Omega_{\mathcal{Y}_1} = \{1, 2, 3, 4, 5, 6\}$
  - $P_{\mathcal{Y}_2}(y_2) = 1/6, \forall y_2 \in \Omega_{\mathcal{Y}_2} = \{1, 2, 3, 4, 5, 6\}$
  - et $P_{\mathcal{Z}}(1) = 1/6$ (les deux mêmes faces sont observées) et $P_{\mathcal{Z}}(0) = 5/6$ (deux faces différentes sont observées).
- En ce qui concerne la relation entre $\mathcal{Y}_1$ et $\mathcal{Y}_2$, on voit que
  - $P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1, y_2) = 1/36, \forall y_1, y_2$
  - et donc $P_{\mathcal{Y}_1,\mathcal{Y}_2} = P_{\mathcal{Y}_1} P_{\mathcal{Y}_2}$, i.e. $\mathcal{Y}_1 \perp \mathcal{Y}_2$.
  - I.e. $P_{\mathcal{Y}_1|\mathcal{Y}_2}(y_1|y_2) = P_{\mathcal{Y}_1}(y_1)$ et $P_{\mathcal{Y}_2|\mathcal{Y}_1}(y_2|y_1) = P_{\mathcal{Y}_2}(y_2)$, $\forall y_1, y_2$.
- En ce qui concerne la relation entre $\mathcal{Y}_1$ et $\mathcal{Z}$, on observe que
  - $P_{\mathcal{Y}_1,\mathcal{Z}}(y_1, 1) = 1/36$ et $P_{\mathcal{Y}_1,\mathcal{Z}}(y_1, 0) = 5/36$, donc $P_{\mathcal{Y}_1,\mathcal{Z}} = P_{\mathcal{Y}_1} P_{\mathcal{Z}}$.
  - Donc $\mathcal{Y}_1 \perp \mathcal{Z}$, et donc aussi $P_{\mathcal{Z}|\mathcal{Y}_1}(z|y_1) = P_{\mathcal{Z}}(z)$.
- On peut évidemment conclure le même genre de choses au sujet de la relation entre $\mathcal{Y}_2$ et $\mathcal{Z}$.

Exemple 2: lois de probabilité induites (b)

NB: la mesure P définie sur Ω est uniforme ($P(\omega) = 1/36, \forall \omega \in \Omega$).

- Pour rappel, on a
  - $P_{\mathcal{X}}(2) = 1/36, P_{\mathcal{X}}(3) = 2/36, \ldots, P_{\mathcal{X}}(11) = 2/36, P_{\mathcal{X}}(12) = 1/36$ ($\mathcal{X} = \mathcal{Y}_1 + \mathcal{Y}_2$).
- On peut se convaincre que
  - $P_{\mathcal{Y}_1,\mathcal{X}}(y_1, x) = 1/36, \forall (y_1, x) : y_1 \in \{1, \ldots, 6\}, x \in \{y_1 + 1, \ldots, y_1 + 6\}$. et 0 ailleurs.
  - Donc $P_{\mathcal{X}|\mathcal{Y}_1}(x|y_1) = 1/6$ si $x \in \{y_1 + 1, \ldots, y_1 + 6\}$, sinon 0; p.ex. $P_{\mathcal{X}|\mathcal{Y}_1}(x|3)$, vaut 0 si $x \in \{2, 3, 10, 11, 12\}$ et 1/6 si $x \in \{4, 5, 6, 7, 8, 9\}$.
  - On a aussi $P_{\mathcal{Y}_1|\mathcal{X}}(y_1|x) = P_{\mathcal{Y}_1,\mathcal{X}}(y_1, x)/P_{\mathcal{X}}(x)$, et par exemple $P_{\mathcal{Y}_1|\mathcal{X}}(3|5) = P_{\mathcal{Y}_1,\mathcal{X}}(3, 5)/P_{\mathcal{X}}(5) = (1/36)/(4/36) = 1/4$, et aussi $P_{\mathcal{Y}_1|\mathcal{X}}(6|5) = P_{\mathcal{Y}_1,\mathcal{X}}(6, 5)/P_{\mathcal{X}}(5) = (0/36)/(4/36) = 0$.
  - On voit que $\mathcal{Y}_1 \not\perp \mathcal{X}$ et on peut transposer que $\mathcal{Y}_2 \not\perp \mathcal{X}$.

Table de contingences pour $P_{\mathcal{Y}_1,\mathcal{X}}$

	$\mathcal{X} =$ 2	3	4	5	6	7	8	9	10	11	12	$\sum_{\Omega_{\mathcal{X}}}$
$\mathcal{Y}_1 = 1$	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	1/6
$\mathcal{Y}_1 = 2$	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	1/6
$\mathcal{Y}_1 = 3$	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	1/6
$\mathcal{Y}_1 = 4$	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	1/6
$\mathcal{Y}_1 = 5$	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	1/6
$\mathcal{Y}_1 = 6$	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$\sum_{\Omega_{\mathcal{Y}_1}}$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

Figure : Table de contingences: valeurs de $P_{\mathcal{Y}_1,\mathcal{X}}$

Table de contingences pour $P_{\mathcal{X}|\mathcal{Y}_1}$

	$\mathcal{X} =$ 2	3	4	5	6	7	8	9	10	11	12	$\sum_{\Omega_\mathcal{X}}$
$\mathcal{Y}_1 = 1$	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	0	0	1
$\mathcal{Y}_1 = 2$	0	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	0	1
$\mathcal{Y}_1 = 3$	0	0	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	1
$\mathcal{Y}_1 = 4$	0	0	0	1/6	1/6	1/6	1/6	1/6	1/6	0	0	1
$\mathcal{Y}_1 = 5$	0	0	0	0	1/6	1/6	1/6	1/6	1/6	1/6	0	1
$\mathcal{Y}_1 = 6$	0	0	0	0	0	1/6	1/6	1/6	1/6	1/6	1/6	1

Figure : Table de contingences: valeurs de $P_{\mathcal{X}|\mathcal{Y}_1}$

Table de contingences pour $P_{\mathcal{Y}_1|\mathcal{X}}$

	$\mathcal{X} =$										
	2	3	4	5	6	7	8	9	10	11	12
$\mathcal{Y}_1 = 1$	1	1/2	1/3	1/4	1/5	1/6	0	0	0	0	0
$\mathcal{Y}_1 = 2$	0	1/2	1/3	1/4	1/5	1/6	1/5	0	0	0	0
$\mathcal{Y}_1 = 3$	0	0	1/3	1/4	1/5	1/6	1/5	1/4	0	0	0
$\mathcal{Y}_1 = 4$	0	0	0	1/4	1/5	1/6	1/5	1/4	1/3	0	0
$\mathcal{Y}_1 = 5$	0	0	0	0	1/5	1/6	1/5	1/4	1/3	1/2	0
$\mathcal{Y}_1 = 6$	0	0	0	0	0	1/6	1/5	1/4	1/3	1/2	1
$\sum_{\Omega_{\mathcal{Y}_1}}$	1	1	1	1	1	1	1	1	1	1	1

Figure : Table de contingences: valeurs de $P_{\mathcal{Y}_1|\mathcal{X}}$

Espérance conditionnelle (a)

Supposons ici que $\mathcal{Y}$ est à valeurs réelles. et considérons alors la variable $\mathcal{Z} = \phi(\mathcal{X})$ définie par

$$z(x) \triangleq \sum_{j=1}^{l} y_j P_{\mathcal{Y}|\mathcal{X}}(y_j|x). \tag{6}$$

Il s'agit d'une variable aléatoire réelle discrète dont le nombre de valeurs différentes est au plus égal au nombre de valeurs différentes de $\mathcal{X}$.

Pour une valeur fixée x_i de $\mathcal{X}$, la valeur $z(x_i)$ est l'espérance de la variable $\mathcal{Y}$ calculée par rapport la loi conditionnelle $P(\cdot|X_i)$ induite sur Ω par l'événement $X_i = \{\omega \in \Omega : \mathcal{X}(\omega) = x_i\}$.

On appelle "espérance conditionnelle" de $\mathcal{Y}$ sachant $\mathcal{X}$, la v.a. :

$$E\{\mathcal{Y}|\mathcal{X}\} \triangleq \sum_{j=1}^{l} y_j P_{\mathcal{Y}|\mathcal{X}}(y_j|\mathcal{X}). \tag{7}$$

Espérance conditionnelle de $\mathcal{X}$ sachant $\mathcal{Y}_1$

	$\mathcal{X}=$ 2	3	4	5	6	7	8	9	10	11	12	$E\{\mathcal{X}\|\mathcal{Y}_1\}$
$\mathcal{Y}_1=1$	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	0	0	4.5
$\mathcal{Y}_1=2$	0	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	0	5.5
$\mathcal{Y}_1=3$	0	0	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	6.5
$\mathcal{Y}_1=4$	0	0	0	1/6	1/6	1/6	1/6	1/6	1/6	0	0	7.5
$\mathcal{Y}_1=5$	0	0	0	0	1/6	1/6	1/6	1/6	1/6	1/6	0	8.5
$\mathcal{Y}_1=6$	0	0	0	0	0	1/6	1/6	1/6	1/6	1/6	1/6	9.5

Figure : Table de contingences: valeurs de $P_{\mathcal{X}|\mathcal{Y}_1}$ et de $E\{\mathcal{X}|\mathcal{Y}_1\}$

Espérance conditionnelle de $\mathcal{Y}_1$ sachant $\mathcal{X}$

	$\mathcal{X} =$										
	2	3	4	5	6	7	8	9	10	11	12
$\mathcal{Y}_1 = 1$	1	1/2	1/3	1/4	1/5	1/6	0	0	0	0	0
$\mathcal{Y}_1 = 2$	0	1/2	1/3	1/4	1/5	1/6	1/5	0	0	0	0
$\mathcal{Y}_1 = 3$	0	0	1/3	1/4	1/5	1/6	1/5	1/4	0	0	0
$\mathcal{Y}_1 = 4$	0	0	0	1/4	1/5	1/6	1/5	1/4	1/3	0	0
$\mathcal{Y}_1 = 5$	0	0	0	0	1/5	1/6	1/5	1/4	1/3	1/2	0
$\mathcal{Y}_1 = 6$	0	0	0	0	0	1/6	1/5	1/4	1/3	1/2	1
$E\{\mathcal{Y}_1 \vert \mathcal{X}\}$	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6

Figure : Table de contingences: valeurs de $P_{\mathcal{Y}_1|\mathcal{X}}$ et de $E\{\mathcal{Y}_1|\mathcal{X}\}$

Espérance conditionnelle (b)

Linéarité de l'espérance conditionnelle.
Soit $\mathcal{Y} = \alpha\mathcal{Y}_1 + \beta\mathcal{Y}_2$ une combinaison linéaire de deux variables aléatoires discrètes finies et réelles, et $\mathcal{X}$ une variable aléatoire discrète finie quelconque, toutes définies sur $(\Omega, \mathcal{E}, P)$. On a

$$E\{\mathcal{Y}|\mathcal{X}\} = \alpha E\{\mathcal{Y}_1|\mathcal{X}\} + \beta E\{\mathcal{Y}_2|\mathcal{X}\}. \tag{8}$$

En effet, cela résulte immédiatement de la linéarité de l'opérateur d'espérance, appliquée pour chacune des valeurs x_i possibles de $\mathcal{X}$ dans le contexte de la loi conditionnelle $P(\cdot|X_i)$.

Espérance conditionnelle (c)

Théorème de l'espérance totale (ou de la moyenne conditionnelle)
Soit $\mathcal{X}$ une variable aléatoire discrète finie quelconque et $\mathcal{Y}$ une variable aléatoire discrète finie à valeurs réelles, définies sur $(\Omega, \mathcal{E}, P)$. On a

$$E\{E\{\mathcal{Y}|\mathcal{X}\}\} = E\{\mathcal{Y}\}. \tag{9}$$

En effet, on peut calculer l'espérance mathématique de la variable $E\{\mathcal{Y}|\mathcal{X}\}$ de la façon suivante :

$$\begin{aligned} E\{E\{\mathcal{Y}|\mathcal{X}\}\} &= \sum_{i=1}^{k} E\{\mathcal{Y}|x_i\}P_{\mathcal{X}}(x_i) = \sum_{i=1}^{k} P_{\mathcal{X}}(x_i) \sum_{j=1}^{l} y_j P_{\mathcal{Y}|\mathcal{X}}(y_j|x_i) \\ &= \sum_{j=1}^{l} y_j \sum_{i=1}^{k} P_{\mathcal{X}}(x_i) P_{\mathcal{Y}|\mathcal{X}}(y_j|x_i) = \sum_{j=1}^{l} y_j P_{\mathcal{Y}}(y_j) = E\{\mathcal{Y}\}. \end{aligned}$$

Calcul de l'espérance de l'espérance conditionnelle de $\mathcal{Y}_1$ sachant $\mathcal{X}$: on trouve l'espérance de $\mathcal{Y}_1$

	$\mathcal{X} =$ 2	3	4	5	6	7	8	9	10	11	12
$\mathcal{Y}_1 = 1$	1	1/2	1/3	1/4	1/5	1/6	0	0	0	0	0
$\mathcal{Y}_1 = 2$	0	1/2	1/3	1/4	1/5	1/6	1/5	0	0	0	0
$\mathcal{Y}_1 = 3$	0	0	1/3	1/4	1/5	1/6	1/5	1/4	0	0	0
$\mathcal{Y}_1 = 4$	0	0	0	1/4	1/5	1/6	1/5	1/4	1/3	0	0
$\mathcal{Y}_1 = 5$	0	0	0	0	1/5	1/6	1/5	1/4	1/3	1/2	0
$\mathcal{Y}_1 = 6$	0	0	0	0	0	1/6	1/5	1/4	1/3	1/2	1
$E\{\mathcal{Y}_1\|\mathcal{X}\}$	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
$P_{\mathcal{X}}$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

$$E\{E\{\mathcal{Y}_1|\mathcal{X}\}\} \stackrel{\triangle}{=} \sum_{x\in\Omega_{\mathcal{X}}} E\{\mathcal{Y}_1|x\}P_{\mathcal{X}}(x) = 3.5 = E\{\mathcal{Y}_1\}$$

Figure : Table de contingences: valeurs de $P_{\mathcal{Y}_1|\mathcal{X}}$ et de $E\{\mathcal{Y}_1|\mathcal{X}\}$

Calcul de l'espérance de l'espérance conditionnelle de $\mathcal{X}$ sachant $\mathcal{Y}_1$: on trouve l'espérance de $\mathcal{X}$

	$\mathcal{X} =$												
	2	3	4	5	6	7	8	9	10	11	12	$E\{\mathcal{X}\|\mathcal{Y}_1\}$	$P_{\mathcal{Y}_1}$
$\mathcal{Y}_1 = 1$	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	0	0	4.5	1/6
$\mathcal{Y}_1 = 2$	0	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	0	5.5	1/6
$\mathcal{Y}_1 = 3$	0	0	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	6.5	1/6
$\mathcal{Y}_1 = 4$	0	0	0	1/6	1/6	1/6	1/6	1/6	1/6	0	0	7.5	1/6
$\mathcal{Y}_1 = 5$	0	0	0	0	1/6	1/6	1/6	1/6	1/6	1/6	0	8.5	1/6
$\mathcal{Y}_1 = 6$	0	0	0	0	0	1/6	1/6	1/6	1/6	1/6	1/6	9.5	1/6
												$E\{\mathcal{X}\} = 7$	

Figure : Table de contingences: valeurs de $P_{\mathcal{X}|\mathcal{Y}_1}$ et de $E\{\mathcal{X}|\mathcal{Y}_1\}$

Espérance conditionnelle (d)

Indépendance $\Rightarrow$ espérance conditionnelle constante

Soit $\mathcal{X}$ une variable aléatoire discrète finie quelconque et $\mathcal{Y}$ une variable aléatoire discrète finie à valeurs réelles, définies sur $(\Omega, \mathcal{E}, P)$.

Si $\mathcal{X}$ et $\mathcal{Y}$ sont indépendantes, alors l'espérance conditionnelle est constante, et $E\{\mathcal{Y}|x\} = E\{\mathcal{Y}\}, \forall x$.

La réciproque de cette propriété n'est pas vraie.

En effet, dans le second membre de l'équation (6), on peut alors remplacer $P_{\mathcal{Y}|\mathcal{X}}(y_j|x)$ par $P_{\mathcal{Y}}(y_j)$, et on obtient bien $z(x_i) = E\{\mathcal{Y}\}, \forall i = 1, \ldots, k$.

Espérance conditionnelle (e)

Espérance conditionnelle d'une fonction de $\mathcal{X}$
Lorsque $\mathcal{Y}$ est une fonction de $\mathcal{X}$, on a $E\{\mathcal{Y}|\mathcal{X}\} = \mathcal{Y}$.

En particulier, on a $E\{\mathcal{Y}|\mathcal{Y}\} = \mathcal{Y}$.

En effet, le second membre de l'équation (6) devient alors pour une valeur donnée x_i,

$$E\{\mathcal{Y}|x_i\} = \sum_{j=1}^{l} y_j P_{\mathcal{Y}|\mathcal{X}}(y_j|x_i) = \sum_{j=1}^{l} y_j \delta_{y_j,\phi(x_i)} = \phi(x_i),$$

où δ est le symbole de Kronecker. Autrement dit, $E\{\mathcal{Y}|x_i\} = \phi(x_i)$, et donc $E\{\mathcal{Y}|\mathcal{X}\} = \phi(\mathcal{X}) = \mathcal{Y}$.

Variance conditionnelle (a)

On définit, la variance conditionnelle de $\mathcal{Y}$ sachant $\mathcal{X}$ comme suit.

Variance conditionnelle

$$V\{\mathcal{Y}|x\} \triangleq \sum_{j=1}^{l} (y_j - E\{\mathcal{Y}|x\})^2 \, P_{\mathcal{Y}|\mathcal{X}}(y_j|x). \tag{10}$$

Comme pour l'espérance conditionnelle, cette formule définit une nouvelle variable aléatoire, notée $V\{\mathcal{Y}|\mathcal{X}\}$, et qui est une fonction de la variable aléatoire $\mathcal{X}$.

Calcul de la variance conditionnelle de $\mathcal{Y}_1$ sachant $\mathcal{X}$

	$\mathcal{X} =$										
	2	3	4	5	6	7	8	9	10	11	12
$\mathcal{Y}_1 = 1$	1	1/2	1/3	1/4	1/5	1/6	0	0	0	0	0
$\mathcal{Y}_1 = 2$	0	1/2	1/3	1/4	1/5	1/6	1/5	0	0	0	0
$\mathcal{Y}_1 = 3$	0	0	1/3	1/4	1/5	1/6	1/5	1/4	0	0	0
$\mathcal{Y}_1 = 4$	0	0	0	1/4	1/5	1/6	1/5	1/4	1/3	0	0
$\mathcal{Y}_1 = 5$	0	0	0	0	1/5	1/6	1/5	1/4	1/3	1/2	0
$\mathcal{Y}_1 = 6$	0	0	0	0	0	1/6	1/5	1/4	1/3	1/2	1
$E\{\mathcal{Y}_1\|\mathcal{X}\}$	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
$V\{\mathcal{Y}_1\|\mathcal{X}\}$	0	3/12	8/12	15/12	24/12	35/12	24/12	15/12	8/12	3/12	0

Figure : Table de contingences: valeurs de $P_{\mathcal{Y}_1|\mathcal{X}}$ et de $V\{\mathcal{Y}_1|\mathcal{X}\}$

Pour rappel: variance d'une v.a. discrète et uniforme sur $\{m, \ldots, m+n-1\} = (n^2-1)/12$.

Variance conditionnelle (b)

Théorème de la variance totale (ou de la variance conditionnelle)

$$V\{\mathcal{Y}\} = E\{V\{\mathcal{Y}|\mathcal{X}\}\} + V\{E\{\mathcal{Y}|\mathcal{X}\}\}. \quad (11)$$

Nous reviendrons sur ce théorème dans sa version générale plus loin dans ce cours.

Indépendance $\Rightarrow$ variance conditionnelle constante

Si $\mathcal{X}$ et $\mathcal{Y}$ sont indépendantes, alors la variance conditionnelle est constante, et $V\{\mathcal{Y}|x\} = V\{\mathcal{Y}\}, \forall x$.
La réciproque n'est pas vraie.

Ce résultat s'obtient directement en remplaçant $E\{\mathcal{Y}|x\}$ par $E\{\mathcal{Y}\}$ et $P_{\mathcal{Y}|\mathcal{X}}(y_j|x)$ par $P_{\mathcal{Y}}(y_j)$ dans (10).

Théorème de la variance Totale pour $\mathcal{Y}_1$ sachant $\mathcal{X}$

	$\mathcal{X} =$										
	2	3	4	5	6	7	8	9	10	11	12
$\mathcal{Y}_1 = 1$	1	1/2	1/3	1/4	1/5	1/6	0	0	0	0	0
$\mathcal{Y}_1 = 2$	0	1/2	1/3	1/4	1/5	1/6	1/5	0	0	0	0
$\mathcal{Y}_1 = 3$	0	0	1/3	1/4	1/5	1/6	1/5	1/4	0	0	0
$\mathcal{Y}_1 = 4$	0	0	0	1/4	1/5	1/6	1/5	1/4	1/3	0	0
$\mathcal{Y}_1 = 5$	0	0	0	0	1/5	1/6	1/5	1/4	1/3	1/2	0
$\mathcal{Y}_1 = 6$	0	0	0	0	0	1/6	1/5	1/4	1/3	1/2	1
$E\{\mathcal{Y}_1 \vert \mathcal{X}\}$	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
$V\{\mathcal{Y}_1 \vert \mathcal{X}\}$	0	3/12	8/12	15/12	24/12	35/12	24/12	15/12	8/12	3/12	0
$P_{\mathcal{X}}$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

$$E\{V\{\mathcal{Y}_1|\mathcal{X}\}\} \triangleq \sum_{x\in\Omega_{\mathcal{X}}} V\{\mathcal{Y}_1|x\}P_{\mathcal{X}}(x) = 630/432$$

On a par ailleurs que $V\{E\{\mathcal{Y}_1|\mathcal{X}\}\} \triangleq \sum_{x\in\Omega_{\mathcal{X}}}(E\{\mathcal{Y}_1|x\} - E\{\mathcal{Y}_1\})^2 P_{\mathcal{X}}(x) = 630/432$

Donc $E\{V\{\mathcal{Y}_1|\mathcal{X}\}\} + V\{E\{\mathcal{Y}_1|\mathcal{X}\}\} = 1260/432 = 35/12 = V\{\mathcal{Y}_1\}$.

Calcul de la variance conditionnelle de $\mathcal{X}$ étant donnée $\mathcal{Y}_1$

- On peut faire un calcul analogue pour vérifier qu'on a aussi $V\{\mathcal{X}\} = E\{V\{\mathcal{X}|\mathcal{Y}_1\}\} + V\{E\{\mathcal{X}|\mathcal{Y}_1\}\}$
- Mais on peut aussi s'en convaincre autrement, en exploitant seulement le fait que $\mathcal{X} = \mathcal{Y}_1 + \mathcal{Y}_2$ et que $\mathcal{Y}_1 \perp \mathcal{Y}_2$.
  On peut en effet calculer que :
  - $V\{\mathcal{X}\} = V\{\mathcal{Y}_1\} + V\{\mathcal{Y}_2\}$ étant donné que $\mathcal{Y}_1 \perp \mathcal{Y}_2$.
  - Ensuite on a aussi $V\{\mathcal{X}|y_1\} = V\{\mathcal{Y}_2\}$ quelque soit la valeur de y_1; par conséquent $E\{V\{\mathcal{X}|\mathcal{Y}_1\}\} = V\{\mathcal{Y}_2\}$.
  - Enfin, $E\{\mathcal{X}|\mathcal{Y}_1\} = \mathcal{Y}_1 + E\{\mathcal{Y}_2\}$, et donc $V\{E\{\mathcal{X}|\mathcal{Y}_1\}\} = V\{\mathcal{Y}_1\}$.
- NB: par symétrie on obtient également $E\{V\{\mathcal{X}|\mathcal{Y}_2\}\} = V\{\mathcal{Y}_1\}$ et $V\{E\{\mathcal{X}|\mathcal{Y}_1\}\} = V\{\mathcal{Y}_2\}$.

Variance conditionnelle (c)

Variance conditionnelle d'une fonction de $\mathcal{X}$

Lorsque $\mathcal{Y}$ est une fonction de $\mathcal{X}$, on a $V\{\mathcal{Y}|\mathcal{X}\} = 0_\Omega$, où nous désignons par 0_Ω une variable aléatoire qui est identiquement nulle sur Ω.

En particulier $V\{\mathcal{Y}|\mathcal{Y}\} = 0_\Omega$.

En effet, le second membre de l'équation (10) devient alors pour une valeur donnée x_i :

$$\begin{aligned}
V\{\mathcal{Y}|x_i\} &\triangleq \sum_{j=1}^{l} (y_j - E\{\mathcal{Y}|x_i\})^2 P_{\mathcal{Y}|\mathcal{X}}(y_j|x_i) \\
&= \sum_{j=1}^{l} (y_j - \phi(x_i))^2 P_{\mathcal{Y}|\mathcal{X}}(y_j|x_i) \\
&= \sum_{j=1}^{l} (y_j - \phi(x_i))^2 \delta_{y_j,\phi(x_i)} = 0.
\end{aligned}$$

Espace $\mathcal{F}_\Omega$ de v.a. réelles sur un espace $(\Omega, \mathcal{E}, P)$ fini

- Soit $\Omega = \{\omega_1, \ldots, \omega_n\}$ (de cardinalité finie $|\Omega| = n$).
- La mesure de probabilité P sur Ω est entièrement définie par les n nombres $P(\omega_i)$ (positifs et de somme égale à 1).
- Une variable aléatoire $\mathcal{X}$ réelle quelconque sur Ω est alors entièrement définie par les n nombres $\mathcal{X}(\omega_i) \in \mathbb{R}$.
- Autrement dit, l'ensemble $\mathcal{F}_\Omega$ des v.a. réelles est en bijection avec $\mathbb{R}^n$.
- Nous définissons une géométrie sur $\mathcal{F}_\Omega$ par l'intermédiaire du

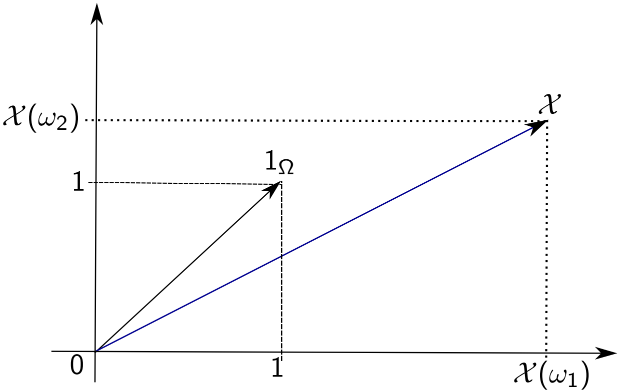
Produit scalaire de deux variables aléatoires

$$\langle \mathcal{X}, \mathcal{Y} \rangle \triangleq \sum_{i=1}^{n} \mathcal{X}(\omega_i)\mathcal{Y}(\omega_i)P(\omega_i). \tag{12}$$

Deux v.a. sont dites **orthogonales** si leur produit scalaire est nul.

Notons qu'on a donc $\langle \mathcal{X}, \mathcal{Y} \rangle = E\{\mathcal{X}\mathcal{Y}\}$.

Représentation graphique de $\mathcal{F}_{\{\omega_1,\omega_2\}}$



NB: pour que $\langle \mathcal{X}, \mathcal{Y} \rangle = 0$ corresponde à l'orthogonalité euclidienne, il faut utiliser des échelles pondérées par $\sqrt{P(\omega_i)}$ pour les axes; ici nous avons supposé que $P(\omega_1) = P(\omega_2) = 0.5$, ce qui correspond à des axes de même échelle (cf. la localisation de 1_Ω).

Propriétés géométriques de l'espace $\mathcal{F}_\Omega$

Norme d'une variable aléatoire

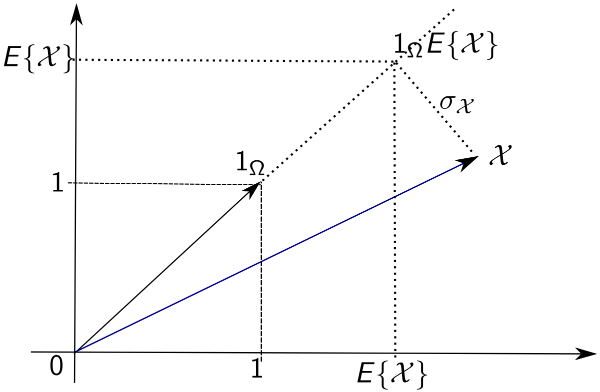
$$\|\mathcal{X}\| \triangleq \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle} = \sqrt{E\{\mathcal{X}^2\}}, \tag{13}$$

Distance de deux variables aléatoires

$$d(\mathcal{X}, \mathcal{Y}) \triangleq \|\mathcal{X} - \mathcal{Y}\| = \sqrt{E\{(\mathcal{X} - \mathcal{Y})^2\}}, \tag{14}$$

- En particulier, deux variables aléatoires sont identiques, si et seulement si leur distance est nulle. On voit également que $\|1_\Omega\| = 1$.
- On a $E\{\mathcal{X}\} = \langle \mathcal{X}, 1_\Omega \rangle$.
- On désigne par $\mathcal{F}_{1_\Omega}$ l'ensemble des v.a. constantes, c'est-à-dire les variables qui sont des produits de 1_Ω par une constante : on dit que $\mathcal{F}_{1_\Omega}$ est la droite des constantes.
- L'ensemble des variables aléatoires de moyenne nulle forme un sous-espace linéaire de $\mathcal{F}_\Omega$ qui est le complément orthogonal de la droite des constantes.

Interprétation géométrique de l'espérance et de la variance



NB: $1_\Omega E\{\mathcal{X}\}$ est la projection de la variable $\mathcal{X}$ sur la droite des constantes. La variable $\mathcal{X} - 1_\Omega E\{\mathcal{X}\}$ est donc orthogonale à la droite des constantes (i.e. de moyenne nulle) et elle est de variance égale $\sigma^2_\mathcal{X}$: on doit donc avoir que $\|\mathcal{X}\|^2 \triangleq E\{(\mathcal{X})^2\} = \sigma^2_\mathcal{X} + (E\{\mathcal{X}\})^2$.

Propriétés géométriques du sous-espace $\mathcal{F}_{\mathcal{X}}$ (a)

Sous-espace $\mathcal{F}_{\mathcal{X}}$

Etant donnée une v.a. $\mathcal{X}$, on désigne par $\mathcal{F}_{\mathcal{X}}$, l'ensemble des v.a. qui peuvent s'écrire sous la forme d'une fonction de la variable $\mathcal{X}$.

- On a $\mathcal{F}_{1_\Omega} \subset \mathcal{F}_{\mathcal{X}}$, et en particulier $0_\Omega \in \mathcal{F}_{\mathcal{X}}$.
- $\mathcal{F}_{\mathcal{X}}$ contient l'ensemble des combinaisons linéaires de ses éléments : on dit que $\mathcal{F}_{\mathcal{X}}$ est un sous-espace linéaire de $\mathcal{F}_\Omega$.
- La v.a. $\mathcal{Z} = E\{\mathcal{Y}|\mathcal{X}\}$ est l'élément de $\mathcal{F}_{\mathcal{X}}$ le plus proche de $\mathcal{Y}$.

  En effet, soit $\phi(\mathcal{X})$ élément quelconque de $\mathcal{F}_{\mathcal{X}}$. On a

  $$E\{(\mathcal{Y} - \phi(\mathcal{X}))^2\} = \sum_{x \in \Omega_{\mathcal{X}}} P_{\mathcal{X}}(x) \sum_{y \in \Omega_{\mathcal{Y}}} (y - \phi(x))^2 P_{\mathcal{Y}|\mathcal{X}}(y|x),$$

  dont la valeur minimale est atteinte en choisissant $\forall x : \phi(x) = E\{\mathcal{Y}|x\} = \sum_{y \in \Omega_{\mathcal{Y}}} y P_{\mathcal{Y}|\mathcal{X}}(y|x)$.
- On dit que $E\{\mathcal{Y}|\mathcal{X}\}$ est la projection orthogonale de $\mathcal{Y}$ sur $\mathcal{F}_{\mathcal{X}}$.
- On peut en déduire le théorème de l'espérance totale : la projection sur 1_Ω peut en effet être calculée en projetant d'abord sur $\mathcal{F}_{\mathcal{X}}$, puis en projetant le résultat sur 1_Ω.

Propriétés géométriques du sous-espace $\mathcal{F}_{\mathcal{X}}$ (b)

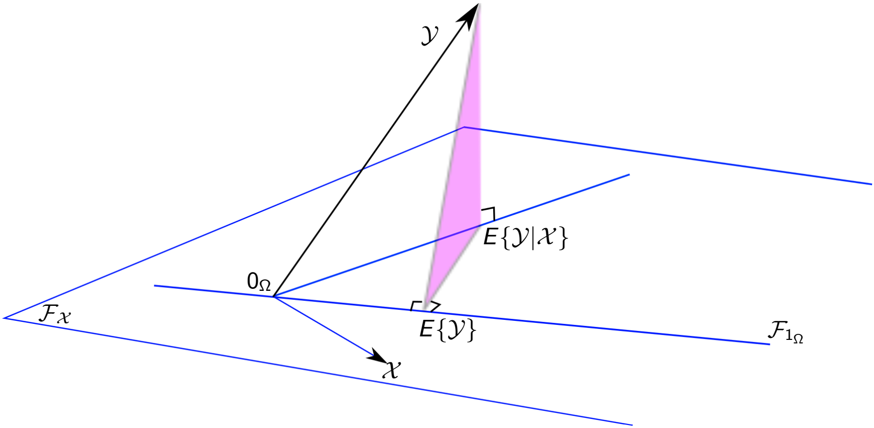
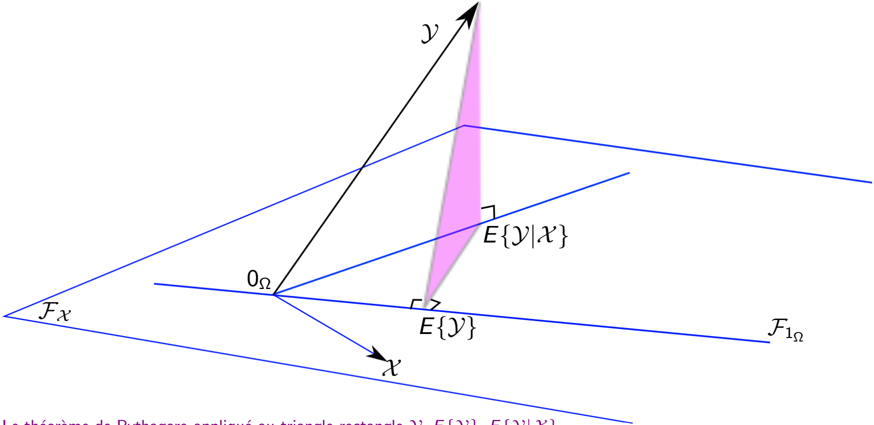


Illustration du théorème de l'espérance totale : on observe "graphiquement" que projeter d'abord $\mathcal{Y}$ sur $\mathcal{F}_{\mathcal{X}}$ (ce qui donne $E\{\mathcal{Y}|\mathcal{X}\}$), et ensuite projeter le résultat $E\{\mathcal{Y}|\mathcal{X}\}$ sur $\mathcal{F}_{1_\Omega}$ (ce qui donne $E\{E\{\mathcal{Y}|\mathcal{X}\}\}$), donne le même résultat que de projeter directement $\mathcal{Y}$ sur $\mathcal{F}_{1_\Omega}$ (ce qui donne $E\{\mathcal{Y}\}$) : $E\{E\{\mathcal{Y}|\mathcal{X}\}\} = E\{\mathcal{Y}\}$.

Propriétés géométriques du sous-espace $\mathcal{F}_{\mathcal{X}}$ (c)



Le théorème de Pythagore appliqué au triangle rectangle $\mathcal{Y}$, $E\{\mathcal{Y}\}$, $E\{\mathcal{Y}|\mathcal{X}\}$ donne $\|\mathcal{Y} - E\{\mathcal{Y}\}\|^2 = \|E\{\mathcal{Y}\} - E\{\mathcal{Y}|\mathcal{X}\}\|^2 + \|\mathcal{Y} - E\{\mathcal{Y}|\mathcal{X}\}\|^2$, i.e. le théorème de la variance totale : $V\{\mathcal{Y}\} = V\{E\{\mathcal{Y}|\mathcal{X}\}\} + \|\mathcal{Y} - E\{\mathcal{Y}|\mathcal{X}\}\|^2$.

En effet $\|\mathcal{Y} - E\{\mathcal{Y}|\mathcal{X}\}\|^2 \triangleq E\{(\mathcal{Y} - E\{\mathcal{Y}|\mathcal{X}\})^2\} = \sum_{x\in\Omega_{\mathcal{X}}} \sum_{y\in\Omega_{\mathcal{Y}}} (y - E\{\mathcal{Y}|x\})^2 P_{\mathcal{X}}(x) P_{\mathcal{Y}|\mathcal{X}}(y|x) = \sum_{x\in\Omega_{\mathcal{X}}} P_{\mathcal{X}}(x) \sum_{y\in\Omega_{\mathcal{Y}}} (y - E\{\mathcal{Y}|x\})^2 P_{\mathcal{Y}|\mathcal{X}}(y|x) = E\{V\{\mathcal{Y}|\mathcal{X}\}\}$.

Orthogonalité ne veut pas dire indépendance !!!!

Pour désigner l'**orthogonalité** de $\mathcal{X}$ et $\mathcal{Y}$ nous écrivons $\langle \mathcal{X}, \mathcal{Y} \rangle = 0$. Nous réservons la notation $\mathcal{X} \perp \mathcal{Y}$ pour indiquer qu'elles sont **indépendantes**.

- Tout élément de $\mathcal{F}_{1_\Omega}$ est indépendant de tout élément de $\mathcal{F}_\Omega$ (y compris de lui-même). Donc, 0_Ω est indépendante de toute variable aléatoire.
- Si $\mathcal{Y} \perp \mathcal{X}$ alors $\forall \mathcal{Z} \in \mathcal{F}_\mathcal{X}$, on a $\mathcal{Y} \perp \mathcal{Z}$.
- Si $\mathcal{Y}$ ou $\mathcal{X}$ est de moyenne nulle, alors $[\mathcal{Y} \perp \mathcal{X} \Rightarrow \langle \mathcal{X}, \mathcal{Y} \rangle = 0]$. Cependant, la réciproque n'est pas vraie.
- $\mathcal{Y} \perp \mathcal{X} \Leftrightarrow (\mathcal{Y} - E\{\mathcal{Y}\}) \perp (\mathcal{X} - E\{\mathcal{X}\})$.
- $\mathcal{Y} \perp \mathcal{X}$, si et seulement si $\langle f(\mathcal{X}), g(\mathcal{Y}) \rangle = 0$ quelles que soient les fonctions f et g de moyennes nulles (voir section 4.3).

On voit donc que la notion d'orthogonalité est intimement liée à la notion d'indépendance mais qu'il s'agit de notions différentes. Ces deux notions sont aussi différentes de celle d'indépendance linéaire de vecteurs dans $\mathbb{R}^n$.

4.2 Variables aléatoires continues et conditionnement

- Une seule des deux variables est continue
- Les deux sont continues
- Cas général
- Problème de régression au sens des moindres carrés

Disons que $\mathcal{Y}$ est continue et que $\mathcal{X}$ est discrète (a)

- On définit la fonction de répartition conditionnelle de $\mathcal{X}$ par

$$F_{\mathcal{Y}|\mathcal{X}}(y|x) \triangleq P(\mathcal{Y} < y|\mathcal{X} = x), \tag{15}$$

dont la dérivé $f_{\mathcal{Y}|\mathcal{X}}(y|x)$ (si définie pp) est la densité conditionnelle.

- La fonction de répartition marginale s'écrit

$$F_{\mathcal{Y}}(y) \triangleq \sum_{i=1}^{k} P_{\mathcal{X}}(x_i) F_{\mathcal{Y}|\mathcal{X}}(y|x_i) \tag{16}$$

qui dérivée terme à terme (si OK) donne la densité marginale

$$f_{\mathcal{Y}}(y) \triangleq \sum_{i=1}^{k} P_{\mathcal{X}}(x_i) f_{\mathcal{Y}|\mathcal{X}}(y|x_i). \tag{17}$$

- Les théorèmes de l'espérance et de la variance totales se transposent.

Disons que $\mathcal{Y}$ est continue et que $\mathcal{X}$ est discrète (b)

- On peut également écrire

$$P(\mathcal{X}=x|\mathcal{Y}<y)=\frac{F_{\mathcal{Y}|\mathcal{X}}(y|x)P_{\mathcal{X}}(x)}{F_{\mathcal{Y}}(y)}, \tag{18}$$

mais nous ne pouvons pas pour le moment écrire

$$P(\mathcal{X}=x|\mathcal{Y}=y)=\frac{f_{\mathcal{Y}|\mathcal{X}}(y|x)P_{\mathcal{X}}(x)}{f_{\mathcal{Y}}(y)}, \tag{19}$$

car la condition $\mathcal{Y}=y$ désigne un événement de probabilité nulle par rapport auquel on ne peut pas en principe conditionner.

Lorsque $\mathcal{X}$ et $\mathcal{Y}$ sont conjointement continues

- Par définition, $\mathcal{X}$ et $\mathcal{Y}$ sont conjointement continues, si $\exists f_{\mathcal{X},\mathcal{Y}}$ (densité conjointe) telle que $\forall a, b \in \mathbb{R} : F_{\mathcal{X},\mathcal{Y}}(a, b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{\mathcal{X},\mathcal{Y}}(x, y)\,\mathrm{d}x\,\mathrm{d}y$.
- Dans ce cas, les deux variables sont continues, et leurs densités marginales et conditionnelles existent aussi.
- En particulier on a (en évitant de diviser par zéro)

$$f_{\mathcal{X}}(x) = \int_{\mathbb{R}} f_{\mathcal{X},\mathcal{Y}}(x, y)\,\mathrm{d}y \qquad \text{et} \qquad f_{\mathcal{Y}}(y) = \int_{\mathbb{R}} f_{\mathcal{X},\mathcal{Y}}(x, y)\,\mathrm{d}x,$$

$$f_{\mathcal{Y}|\mathcal{X}}(y|x) = \frac{f_{\mathcal{X},\mathcal{Y}}(x, y)}{f_{\mathcal{X}}(x)} \qquad \text{et} \qquad f_{\mathcal{X}|\mathcal{Y}}(x|y) = \frac{f_{\mathcal{X},\mathcal{Y}}(x, y)}{f_{\mathcal{Y}}(y)},$$

$$E\{\mathcal{Y}|x\}=\int_{\mathbb{R}} y f_{\mathcal{Y}|\mathcal{X}}(y|x)\,\mathrm{d}y \quad \text{et} \quad V\{\mathcal{Y}|x\}=\int_{\mathbb{R}} (y - E\{\mathcal{Y}|x\})^2 f_{\mathcal{Y}|\mathcal{X}}(y|x)\,\mathrm{d}y,$$

$$E\{\mathcal{X}|y\}=\int_{\mathbb{R}} x f_{\mathcal{X}|\mathcal{Y}}(x|y)\,\mathrm{d}x \quad \text{et} \quad V\{\mathcal{X}|y\}=\int_{\mathbb{R}} (x - E\{\mathcal{X}|y\})^2 f_{\mathcal{X}|\mathcal{Y}}(x|y)\,\mathrm{d}x,$$

- et les formules de Bayes

$$f_{\mathcal{Y}|\mathcal{X}}(y|x) = \frac{f_{\mathcal{X}|\mathcal{Y}}(x|y) f_{\mathcal{Y}}(y)}{f_{\mathcal{X}}(x)} \quad \text{et} \quad f_{\mathcal{X}|\mathcal{Y}}(x|y) = \frac{f_{\mathcal{Y}|\mathcal{X}}(y|x) f_{\mathcal{X}}(x)}{f_{\mathcal{Y}}(y)}.$$

Cas général

- La définition générale rigoureuse de la notion de mesure de probabilité conditionnelle, et des notions d'espérance et de variances conditionnelles associées, sort du cadre de ce cours introductif.
- Dans le cadre de ce cours, on peut se contenter de l'information suivante :
  - si $\mathcal{Y}$ est une variable aléatoire réelle, et si $\mathcal{X}$ est une variable aléatoire soit discrète, soit à valeurs dans $\mathbb{R}^p$, alors il est permis de conditionner Ω et donc $\mathcal{Y}$ par rapport à $\mathcal{X}$ localement.
  - De plus, si $E\{\mathcal{Y}\}$ existe alors il existe une v.a. aléatoire "espérance conditionnelle" qui satisfait au théorème de l'espérance totale. Enfin, si $V\{\mathcal{Y}\}$ existe, alors la variance conditionnelle existe aussi, et cette v.a. satisfait alors aussi au théorème de la variance totale.

Régression linéaire au sens des moindres carrés (a)

- Position générale du problème : (traité au chapitre 5)
  - Soit une v.a. cible $\mathcal{Y}$ et soient $\mathcal{X}_1, \ldots, \mathcal{X}_n$ des v.a.*explicatives*.
  - On cherche une fonction affine $\mathcal{Z} = a + \sum_{i=1}^{n} b_i \mathcal{X}_i$ telle que

$$\|\mathcal{Y} - \mathcal{Z}\|^2 = E\left\{\left(\mathcal{Y} - (a + \sum_{i=1}^{n} b_i \mathcal{X}_i)\right)^2\right\}$$

    soit minimale.
- Nous considérons le cas particulier suivant :
  - Une seule variable explicative $\mathcal{X}$.
  - $\mathcal{X}$ et $\mathcal{Y}$ de variances finies et non-nulles
  - $\mathcal{X}$ et $\mathcal{Y}$ conjointement continues de densité $f_{\mathcal{X},\mathcal{Y}}$.
  - On cherche donc à déterminer $a, b \in \mathbb{R}$ tels que

$$MSE(a,b) \triangleq \|\mathcal{Y} - (a + b\mathcal{X})\|^2 = \int_{\mathbb{R}}\int_{\mathbb{R}} (y - a - bx)^2 f_{\mathcal{X},\mathcal{Y}}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

    soit minimale.

Régression linéaire au sens des moindres carrés (b)

- Remarques en ce qui concerne la minimisation de $MSE(a,b)$
  - $MSE(a,b)$ est bien défini $\forall a, b$ puisque les v.a. $\mathcal{X}$ et $\mathcal{Y}$ sont de variances finies, donc aussi toute v.a. qui s'exprime comme une combinaison linéaire de ces variables.
  - $MSE(a,b) \geq 0, \forall a, b$ est continûment dérivable par rapport à a et b. Elle doit donc posséder un minimum global.
- Conditions nécessaires d'optimalité
  - $\frac{\partial MSE(a,b)}{\partial a} = 0, \frac{\partial MSE(a,b)}{\partial b} = 0.$
  - On tire, après quelques étapes (voir 4.2.3.1)

$$b = \frac{E\{\mathcal{X}\mathcal{Y}\} - E\{\mathcal{X}\}E\{\mathcal{Y}\}}{E\{\mathcal{X}^2\} - (E\{\mathcal{X}\})^2} = \frac{\mathrm{cov}\{\mathcal{X};\mathcal{Y}\}}{V\{\mathcal{X}\}}, \tag{20}$$

$$a = E\{\mathcal{Y}\} - bE\{\mathcal{X}\} = E\{\mathcal{Y}\} - \frac{\mathrm{cov}\{\mathcal{X};\mathcal{Y}\}}{V\{\mathcal{X}\}}E\{\mathcal{X}\}. \tag{21}$$

  - Pour rappel : $\mathrm{cov}\{\mathcal{X};\mathcal{Y}\} = E\{(\mathcal{X} - E\{\mathcal{X}\})(\mathcal{Y} - E\{\mathcal{Y}\})\} = E\{\mathcal{X}\mathcal{Y}\} - E\{\mathcal{X}\}E\{\mathcal{Y}\}$, et $\rho_{\mathcal{X};\mathcal{Y}} \triangleq \frac{\mathrm{cov}\{\mathcal{X};\mathcal{Y}\}}{\sigma_{\mathcal{X}}\sigma_{\mathcal{Y}}}$.

Régression linéaire au sens des moindres carrés (c)

- La v.a. $\mathcal{Z}$ définie par

$$\mathcal{Z} = \mu_{\mathcal{Y}} + \rho_{\mathcal{X};\mathcal{Y}}\sigma_{\mathcal{Y}}\frac{(\mathcal{X} - \mu_{\mathcal{X}})}{\sigma_{\mathcal{X}}} \tag{22}$$

  est donc la meilleure approximation au sens des moindres carrés de la v.a. $\mathcal{Y}$ par une fonction affine de la v.a. $\mathcal{X}$.
- $\rho_{\mathcal{X};\mathcal{Y}}$ est appelé pour cette raison coefficient de corrélation linéaire.
- La formule (22) reste valide, même si $\mathcal{X}$ et $\mathcal{Y}$ ne sont pas conjointement continues, pour autant que leurs variances soient > 0.
- Géométriquement :
  - $\mathcal{Z}$ est la projection orthogonale de la v.a. $\mathcal{Y}$ sur le sous-espace de v.a. qui sont des fonctions affines de $\mathcal{X}$.
  - Si $E\{\mathcal{Y}|\mathcal{X}\}$ est une fonction affine de $\mathcal{X}$ alors on doit donc avoir $\mathcal{Z} = E\{\mathcal{Y}|\mathcal{X}\}$.
- Au chapitre 5, on étudiera le problème général (avec plusieurs variables explicatives) dans le cas où les variables sont conjointement gaussiennes.

Exemple : régression linéaire de $\mathcal{Y}_1$ en fonction de $\mathcal{X}$

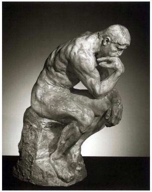
	$\mathcal{X} =$										
	2	3	4	5	6	7	8	9	10	11	12
$\mathcal{Y}_1 = 1$	1	1/2	1/3	1/4	1/5	1/6	0	0	0	0	0
$\mathcal{Y}_1 = 2$	0	1/2	1/3	1/4	1/5	1/6	1/5	0	0	0	0
$\mathcal{Y}_1 = 3$	0	0	1/3	1/4	1/5	1/6	1/5	1/4	0	0	0
$\mathcal{Y}_1 = 4$	0	0	0	1/4	1/5	1/6	1/5	1/4	1/3	0	0
$\mathcal{Y}_1 = 5$	0	0	0	0	1/5	1/6	1/5	1/4	1/3	1/2	0
$\mathcal{Y}_1 = 6$	0	0	0	0	0	1/6	1/5	1/4	1/3	1/2	1
$E\{\mathcal{Y}_1 \vert \mathcal{X}\}$	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6

- On observe que $E\{\mathcal{Y}_1|\mathcal{X}\} = 0.5\mathcal{X}$ et est donc une fonction affine de $\mathcal{X}$.
- Donc, la régression linéaire de $\mathcal{Y}_1$ par $a + b\mathcal{X}$ correspond à $a = 0$ et $b = 0.5$.
- Puisque $b = 0.5 = \mathrm{cov}\{\mathcal{X};\mathcal{Y}_1\}/V\{\mathcal{X}\}$, on a $\mathrm{cov}\{\mathcal{X};\mathcal{Y}_1\} = 35/12 = V\{\mathcal{Y}_1\}$.
  Pour rappel, on a $V\{\mathcal{Y}_1\} = 35/12$, $V\{\mathcal{X}\} = 2V\{\mathcal{Y}_1\} = 70/12$.
- On en déduit que $\rho_{\mathcal{X};\mathcal{Y}_1} = \frac{\mathrm{cov}\{\mathcal{X};\mathcal{Y}_1\}}{\sigma_{\mathcal{X}}\sigma_{\mathcal{Y}_1}} = \frac{V\{\mathcal{Y}_1\}}{\sqrt{V\{\mathcal{Y}_1\}}\sqrt{V\{\mathcal{X}\}}} = 1/\sqrt{2}$

Homework : poursuivre cet exemple

- Trouver la régression linéaire de $\mathcal{Y}_2$ en fonction $\mathcal{X}$ (p.ex. par analogie), et déterminer la valeur de $\rho_{\mathcal{X};\mathcal{Y}_2}$.
- Trouver la régression linéaire de $\mathcal{X}$ en fonction de $\mathcal{Y}_1$ et $\rho_{\mathcal{Y}_1;\mathcal{X}}$.
- Trouver la régression linéaire de $\mathcal{Y}_1$ en fonction de $\mathcal{Y}_2$ et $\rho_{\mathcal{Y}_1;\mathcal{Y}_2}$.
- Trouver les coefficients a, b_1 et b_2 tels que la v.a. $\mathcal{Z} = a + b_1\mathcal{Y}_1 + b_2\mathcal{Y}_2$ soit la plus proche possible de $\mathcal{X}$ au sens des moindres carrés.
  - Quelle est la fonction $\phi(\mathcal{Y}_1, \mathcal{Y}_2)$ la plus proche de $\mathcal{X}$ au sens de la norme induite par le produit scalaire défini sur l'ensemble de v.a. définies sur l'espace $(\Omega, \mathcal{E}, P)$ du double lancer de dés ?
  - Comment définiriez-vous l'espérance conditionnelle $E\{\mathcal{X}|\mathcal{Y}_1, \mathcal{Y}_2\}$, et quelle serait alors sa valeur ?

• 4.3 Synthèse géométrique du problème de régression



Homework pour après les vacances :
lire absolument la section 4.3

Démarche formelle et résultats principaux

- Dans ce qui suit nous esquissons une démarche formelle qui permet d'étudier de façon rigoureuse l'ensemble des variables aléatoires à valeurs réelles sur un espace de probabilité quelconque.
- Notre premier but est de vous illustrer comment on fait pour structurer l'étude mathématique de notions aussi complexes que des ensembles de variables aléatoires.
- Ensuite nous insistons sur quelques résultats importants qui peuvent être établis via cette démarche.

4.3.1 Espace de v.a. à valeurs réelles

- Partant de $(\Omega, \mathcal{E}, P)$, on définit l'ensemble $\mathcal{F}_\Omega$ comme suit.

$\mathcal{F}_\Omega$ est l'ensemble des fonctions $(\mathcal{E}, \mathcal{B}_\mathbb{R})$ mesurables.

Ici $\mathcal{B}_\mathbb{R}$ désigne l'ensemble des parties "boréliennes" de $\mathbb{R}$...

- Dans $\mathcal{F}_\Omega$ on définit la notion "d'égalité presque sûrement" par

Egalité "presque sûrement"

$$\mathcal{X} \overset{p.s.}{=} \mathcal{Y} \Leftrightarrow P(\mathcal{X}(\omega) \neq \mathcal{Y}(\omega)) = 0. \tag{23}$$

Cette relation est une relation d'équivalence, c'est-à-dire qu'elle est

- partout définie sur $\mathcal{F}_\Omega \times \mathcal{F}_\Omega$,
- réflexive : $\forall \mathcal{X} \in \mathcal{F}_\Omega : \left(\mathcal{X} \overset{p.s.}{=} \mathcal{X}\right)$,
- symétrique : $\forall \mathcal{X}, \mathcal{Y} \in \mathcal{F}_\Omega : \left(\mathcal{X} \overset{p.s.}{=} \mathcal{Y}\right) \Rightarrow \left(\mathcal{Y} \overset{p.s.}{=} \mathcal{X}\right)$,
- et transitive. $\forall \mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{F}_\Omega : \left[\left(\mathcal{X} \overset{p.s.}{=} \mathcal{Y}\right) \wedge \left(\mathcal{Y} \overset{p.s.}{=} \mathcal{Z}\right)\right] \Rightarrow \left(\mathcal{X} \overset{p.s.}{=} \mathcal{Z}\right)$.

- On considère les classes d'équivalence de v.a. p.s. égales.

4.3.2 Espace de Hilbert L^2_Ω de v.a. de carré intégrable

L'espace linéaire L^2_Ω est l'ensemble (des classes d'équivalence induites par la relation $\stackrel{p.s.}{=}$) des variables aléatoires de carré intégrable définies sur $(\Omega, \mathcal{E}, P)$.

- $\mathcal{X} \in L^2_\Omega$ veut dire que $E\{\mathcal{X}^2\} < \infty$.
- NB: si $\mathcal{X} \stackrel{p.s.}{=} \mathcal{Y}$, alors $\mathcal{X}^2 \stackrel{p.s.}{=} \mathcal{Y}^2$, et $E\{\mathcal{X}^2\} = E\{\mathcal{Y}^2\}$.
- Si $\mathcal{X}, \mathcal{Y} \in L^2_\Omega$, et $c \in \mathbb{R}$ alors $c\mathcal{X} \in L^2_\Omega$, $\mathcal{X} + \mathcal{Y} \in L^2_\Omega$, $E\{|\mathcal{X}\mathcal{Y}|\} < \infty$, $E\{|\mathcal{X}|\} < \infty$
  Démonstration :
  (a) On a évidemment que $E\{(c\mathcal{X})^2\} = c^2 E\{\mathcal{X}^2\} < \infty$.
  (b) Comme $0 \leq (\mathcal{X} + \mathcal{Y})^2 = 2(\mathcal{X}^2 + \mathcal{Y}^2) - (\mathcal{X} - \mathcal{Y})^2 \leq 2(\mathcal{X}^2 + \mathcal{Y}^2)$, on a donc forcément que $\mathcal{X} + \mathcal{Y} \in L^2_\Omega$.
  (c) Comme $0 \leq (|\mathcal{X}| - |\mathcal{Y}|)^2 = \mathcal{X}^2 + \mathcal{Y}^2 - 2|\mathcal{X}||\mathcal{Y}|$, on a donc $|\mathcal{X}||\mathcal{Y}| \leq \frac{1}{2}(\mathcal{X}^2 + \mathcal{Y}^2)$ et par conséquent $E\{|\mathcal{X}||\mathcal{Y}|\} < \infty$.
  (d) Comme $1_\Omega \in L^2_\Omega$, on a donc aussi que $E\{|\mathcal{X}||1_\Omega|\} = E\{|\mathcal{X}|\} < \infty$. Donc $\mathcal{X}$ doit aussi être d'espérance finie.
- L'ensemble L^2_Ω est donc bien un *sous-espace linéaire* de l'espace de variables aléatoires définies sur Ω. Nous restreignons la suite de notre discussion à ce sous-espace.
- Le **vecteur nul** de L^2_Ω correspond à l'ensemble de variables aléatoires presque sûrement nulles; il est désigné par 0_Ω. On a évidemment $E\{(0_\Omega)^2\} = E\{0_\Omega\} = 0$, et $\forall \mathcal{X} \in L^2_\Omega : \mathcal{X} = \mathcal{X} + 0_\Omega$.
- On désigne par $L^2_{1_\Omega}$ le sous-espace linéaire de L^2_Ω composé des v.a. constantes (p.s. égales à une constante).

Graphiquement ....

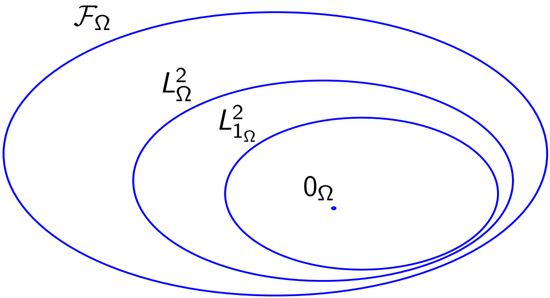


Figure : Relations entre les sous-ensembles de variables aléatoires $\mathcal{F}_\Omega, L^2_\Omega, L^2_{1_\Omega}$ et l'élément nul 0_Ω.

NB: on a démontré au slide précédent que $L^2_\Omega \subset L^1_\Omega$: variance finie implique espérance finie !

Sous-espaces $L^2_{\mathcal{X}}$ et $L^2_{\text{aff}(\mathcal{X})}$

Si $\mathcal{X} \in L^2_\Omega$, nous désignons par $L^2_{\mathcal{X}}$ le sous-ensemble de L^2_Ω de v.a. p.s. égales à une fonction de $\mathcal{X}$. Il contient la droite des constantes.

- $L^2_{\mathcal{X}}$ est un sous-espace linéaire de L^2_Ω, car il contient toutes ses combinaisons linéaires, puisque la combinaison linéaire de fonctions de carré intégrable de $\mathcal{X}$ est aussi une fonction de carré intégrable de $\mathcal{X}$.
- En particulier, $L^2_{\mathcal{X}}$ contient (quelle que soit $\mathcal{X} \in L^2_\Omega$) toutes les variables aléatoires constantes (puisque toute variable aléatoire constante peut être décrite comme une fonction de $\mathcal{X}$). On a donc $L^2_{1_\Omega} \subset L^2_{\mathcal{X}}$.
- Si $\mathcal{X}$ est une variable aléatoire constante, on a $L^2_{\mathcal{X}} = L^2_{1_\Omega}$, qui est de dimension 1.

  Plus généralement, si $\mathcal{X}$ est une variable aléatoire discrète prenant un nombre fini (disons m) de valeurs différentes sur Ω, alors $L^2_{\mathcal{X}}$ est de dimension finie égale à m.

  Dans tous les autres cas, $L^2_{\mathcal{X}}$ est de dimension infinie.

  La dimension de $L^2_{\mathcal{X}}$ résume la richesse de l'information fournie par $\mathcal{X}$ sur ω. Par exemple, si $\mathcal{Y}$ est une fonction de $\mathcal{X}$ alors la dimension de $L^2_{\mathcal{Y}}$ est au plus égale à la dimension de $L^2_{\mathcal{X}}$.

L'ensemble de variables aléatoires qui peuvent s'écrire sous la forme d'une fonction affine d'une variable $\mathcal{X}$ (i.e. sous la forme $\mathcal{Y} = a + b\mathcal{X}$, avec $a, b \in \mathbb{R}$) est aussi un sous-espace linéaire de L^2_Ω. Nous notons ce sous-espace par $L^2_{\text{aff}(\mathcal{X})}$. Il contient la droite des constantes et est inclus dans $L^2_{\mathcal{X}}$.

Graphiquement ....

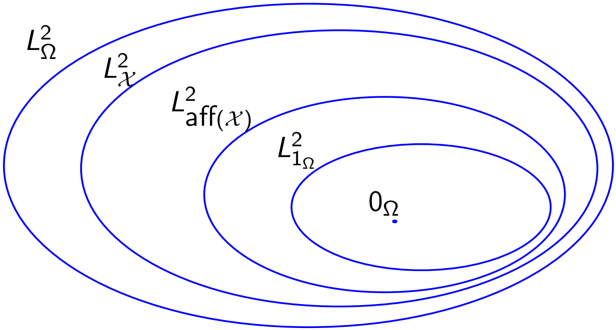


Figure : Relations entre les sous-ensembles de variables aléatoires $L^2_\Omega, L^2_\mathcal{X}, L^2_{\mathrm{aff}(\mathcal{X})}, L^2_{1_\Omega}$ et l'élément nul 0_Ω.

L'espace $L^2_{\mathrm{aff}(\mathcal{X})}$ est un plan ou une droite (dim 2 ou 1)

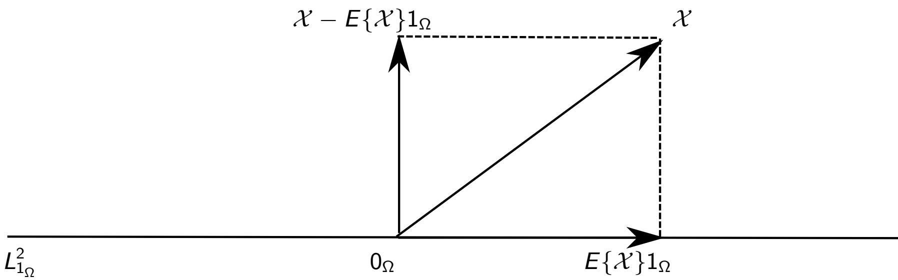


Figure : Le sous-espace linéaire engendré par les v.a. 1_Ω et $\mathcal{X}$ constitue l'espace $L^2_{\mathrm{aff}(\mathcal{X})}$. Si $V\{\mathcal{X}\} = 0$, $\mathcal{X} \overset{p.s.}{=} 1_\Omega E\{\mathcal{X}\}$ et $L^2_{\mathrm{aff}(\mathcal{X})}$ se réduit à $L^2_{1_\Omega}$

Géométrie de L^2_Ω : produit scalaire et orthogonalité

On définit le **produit scalaire** entre deux variables aléatoires par

$$\langle \mathcal{X}, \mathcal{Y} \rangle \triangleq E\{\mathcal{X}\mathcal{Y}\}. \tag{24}$$

- Cette quantité est bien définie (et finie) sur L^2_Ω, et vérifie **l'inégalité de Cauchy-Schwarz**

$$\langle \mathcal{X}, \mathcal{Y} \rangle^2 \leq \langle \mathcal{X}, \mathcal{X} \rangle \langle \mathcal{Y}, \mathcal{Y} \rangle, \tag{25}$$

  l'égalité étant vérifiée si et seulement si $\mathcal{X}$ et $\mathcal{Y}$ sont linéairement dépendantes, c'est-à-dire si l'une des deux peut s'écrire sous la forme d'un multiple de l'autre.
- Le produit scalaire ainsi défini, est une forme symétrique (i.e. $\langle \mathcal{X}, \mathcal{Y} \rangle = \langle \mathcal{Y}, \mathcal{X} \rangle$) et bi-linéaire, i.e.

$$\langle \mathcal{X}, \alpha\mathcal{Y}_1 + \beta\mathcal{Y}_2 \rangle = \alpha\langle \mathcal{X}, \mathcal{Y}_1 \rangle + \beta\langle \mathcal{X}, \mathcal{Y}_2 \rangle. \tag{26}$$

$$\langle \alpha\mathcal{X}_1 + \beta\mathcal{X}_2, \mathcal{Y} \rangle = \alpha\langle \mathcal{X}_1, \mathcal{Y} \rangle + \beta\langle \mathcal{X}_2, \mathcal{Y} \rangle. \tag{27}$$

Deux éléments de L^2_Ω sont **orthogonaux** si leur produit scalaire vaut zéro.

Topologie de L^2_Ω : norme et distance

- Le produit scalaire $\langle \cdot, \cdot \rangle$ induit la notion de norme sur L^2_Ω par

$$\|\mathcal{X}\| = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle} = \sqrt{E\{\mathcal{X}^2\}} = \sqrt{\sigma^2_{\mathcal{X}} + \mu^2_{\mathcal{X}}}. \tag{28}$$

  - On a $\|\mathcal{X}\| \geq 0$, l'égalité étant obtenue si et seulement si $\mathcal{X} \stackrel{p.s.}{=} 0_\Omega$.
  - $\|1_\Omega\| = 1$, et $\|0_\Omega\| = 0$.
  - Sur la Figure du slide 58 on voit que le carré de l'hypoténuse du triangle rectangle formé par 0_Ω, $\mathcal{X}$ et $E\{\mathcal{X}\}1_\Omega$ vaut la somme des carrés de ses côtés, c'est-à-dire $E\{\mathcal{X}^2\} = \mu^2_{\mathcal{X}} + \sigma^2_{\mathcal{X}}$.

- La norme $\|\cdot\|$ induit la notion de distance entre deux éléments de L^2_Ω par

$$d(\mathcal{X}, \mathcal{Y}) = \|\mathcal{X} - \mathcal{Y}\|. \tag{29}$$

Cette distance vérifie l'inégalité triangulaire : $d(\mathcal{X}, \mathcal{Y}) \leq d(\mathcal{X}, \mathcal{Z}) + d(\mathcal{Z}, \mathcal{Y})$.

On a, en effet, $\|\mathcal{X} - \mathcal{Z} + \mathcal{Z} - \mathcal{Y}\|^2 = \|\mathcal{X} - \mathcal{Z}\|^2 + 2\langle \mathcal{X} - \mathcal{Z}, \mathcal{Z} - \mathcal{Y} \rangle + \|\mathcal{Z} - \mathcal{Y}\|^2$ et l'inégalité de Cauchy-Schwarz nous dit que $\langle \mathcal{X} - \mathcal{Z}, \mathcal{Z} - \mathcal{Y} \rangle \leq \|X - Z\| \|Z - Y\| = d(\mathcal{X}, \mathcal{Z}) d(\mathcal{Z}, \mathcal{Y})$.

Topologie de L^2_Ω : ouverts, fermés, et convergence de suites

- La notion de distance induit une topologie sur L^2_Ω, par la notion de boule ouverte centrée sur un élément quelconque de L^2_Ω qui permet à son tour de définir la notion de sous-ensemble ouvert L^2_Ω (un ensemble qui contient une boule ouverte centrée autour de chacun de ses éléments) et de sous-ensemble fermé de L^2_Ω (ceux qui sont les complémentaires dans L^2_Ω d'un sous-ensemble ouvert). La notion de distance permet aussi de définir la notion de convergence des suites d'éléments de L^2_Ω.
- On démontre que dans L^2_Ω toute suite de variables aléatoires qui est de Cauchy, converge vers une variable aléatoire de L^2_Ω (une suite de Cauchy est une suite dont les éléments successifs deviennent de plus en plus proches). Un espace qui possède cette propriété est appelé "espace complet". Si cet espace complet est un espace linéaire normé on dit que c'est un espace de Banach. Si la norme de l'espace de Banach est induite par un produit scalaire, on dit que c'est un **espace de Hilbert**.
- L^2_Ω est donc un espace de Hilbert. Les propriétés des éléments des espaces de Hilbert sont le fruit de l'étude conjointe des notions de combinaison linéaire, de projection orthogonale, et de convergence.
- De façon essentielle, le fait que l'espace L^2_Ω est un espace de Hilbert implique que si la suite $\mathcal{X}_n$ de variables aléatoires de L^2_Ω est telle que la série des normes converge, i.e.

$$\sum_{k=1}^{\infty} \|\mathcal{X}_k\| < \infty, \tag{30}$$

alors la suite des sommes partielles définie par

$$\mathcal{Y}_n \triangleq \sum_{k=1}^{n} \mathcal{X}_k \tag{31}$$

converge vers un élément de L^2_Ω.

Corrélation linéaire vs. orthogonalité

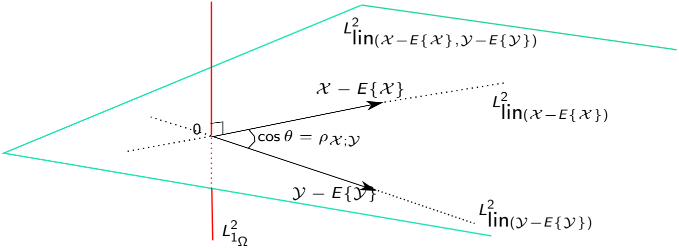
Cauchy-Schwarz appliqué aux variables $\mathcal{X} - E\{\mathcal{X}\}$ et $\mathcal{Y} - E\{\mathcal{Y}\}$ donne

$$\langle \mathcal{X} - E\{\mathcal{X}\}, \mathcal{Y} - E\{\mathcal{Y}\}\rangle^2 \leq \|\mathcal{X} - E\{\mathcal{X}\}\|^2 \|\mathcal{Y} - E\{\mathcal{Y}\}\|^2 \quad (32)$$

ou encore
$$|\mathrm{cov}\{\mathcal{X};\mathcal{Y}\}| \leq \sigma_{\mathcal{X}}\sigma_{\mathcal{Y}} \quad (33)$$

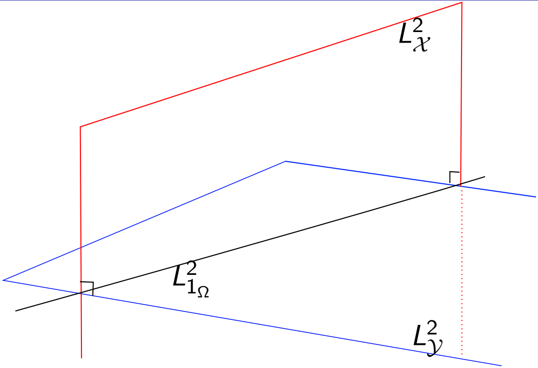
i.e.
$$|\rho_{\mathcal{X};\mathcal{Y}}| \leq 1 \quad (34)$$

$\mathcal{X}$ et $\mathcal{Y}$ sont linéairement non-corrélées si $\rho_{\mathcal{X};\mathcal{Y}} = 0$, i.e. si $\mathcal{X} - E\{\mathcal{X}\}$ et $\mathcal{Y} - E\{\mathcal{Y}\}$ sont orthogonales.



$\rho_{\mathcal{X};\mathcal{Y}}$ est le cosinus de l'angle formé par $\mathcal{X} - E\{\mathcal{X}\}$ et $\mathcal{Y} - E\{\mathcal{Y}\}$. Le plan $L^2_{\mathrm{lin}(\mathcal{X}-E\{\mathcal{X}\},\mathcal{Y}-E\{\mathcal{Y}\})}$ est orthogonal à $L^2_{1_\Omega}$ et représente l'ensemble des v.a. $\mathcal{Z}$ de forme $\mathcal{Z} = \alpha(\mathcal{X} - E\{\mathcal{X}\}) + \beta(\mathcal{Y} - E\{\mathcal{Y}\})$.

Orthogonalité vs. indépendance



Les variables $\mathcal{X}$ et $\mathcal{Y}$ sont indépendantes si et seulement si les espaces de fonctions $L^2_{\mathcal{X}}$ et $L^2_{\mathcal{Y}}$ sont "orthogonaux le long de la droite des constantes $L^2_{1_\Omega}$". $L^2_{\mathcal{X}} \cap L^2_{\mathcal{Y}}$ est alors la droite des constantes $L^2_{1_\Omega}$ et la projection d'un point de $L^2_{\mathcal{X}}$ sur $L^2_{\mathcal{Y}}$ appartient alors à $L^2_{1_\Omega}$ de même que la projection d'un point de $L^2_{\mathcal{Y}}$ sur $L^2_{\mathcal{X}}$. La projection d'un point $\mathcal{Z}$ de $L^2_{\mathcal{X}}$ sur $L^2_{\mathcal{Y}}$ (et aussi d'un point $\mathcal{V}$ de $L^2_{\mathcal{Y}}$ sur $L^2_{\mathcal{X}}$) est donc une v.a. (constante) appartenant à $L^2_{1_\Omega}$ et la variable $\mathcal{Z} - E\{\mathcal{Z}\}$ est orthogonale à $L^2_{\mathcal{Y}}$. Pour tout couple $(\mathcal{Z}, \mathcal{V}) \in L^2_{\mathcal{X}} \times L^2_{\mathcal{Y}}$ on a donc $\mathrm{cov}\{\mathcal{Z}; \mathcal{V}\} = 0$

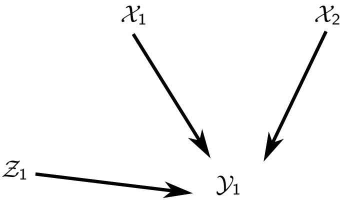
Motivations

- La démarche théorique construit le modèle probabiliste comme suit :
  1. On définit Ω, on lui associe un $\mathcal{E}$ naturel, et on postule P.
  2. On définit des variables aléatoires $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \ldots$ sur Ω
  3. On déduit lois conjointes/conditionnelles de sous-ensembles de v.a.
  4. On déduit éventuellement des relations d'indépendance
  5. On exploite ces informations pour approximer certaines v.a. en fonction d'autres v.a.
- En pratique on procède plutôt comme suit :
  1. on part d'une liste de v.a. d'intérêt pour le problème étudié
  2. on en postule certaines propriétés (des lois, des indépendances)
  3. on en déduit la loi conjointe des v.a., les lois conditionnelles, etc.
  4. ce faisant, on postule qu'il existe un espace $(\Omega, \mathcal{E}, P)$ compatible avec ces hypothèses, sans le définir de manière explicite.
- L'objectif de cette partie du cours est de "joindre les deux bouts" en montrant comment la démarche pratique peut être mise en oeuvre systématiquement pour construire un modèle probabiliste cohérent.

4.4.1 Illustration sur base du double p.o.f. bruité

- Nous allons illustrer en parallèle les deux approches, pour un problème pratique de diagnostic.
- Problème de diagnostic "canonique"
  - On définit un ensemble de v.a. $\mathcal{X}_i$ qui représentent l'état du système (batterie, démarreur, réservoir, etc.); ces variables ne sont pas directement observables.
  - On définit un ensemble de v.a. $\mathcal{Y}_j$ qui représentent les informations observables (radio fonctionne, voiture démarre, plein d'essence, etc.)
  - On modélise les relations entre les variables $\mathcal{X}_i$ et les variables $\mathcal{Y}_j$.
  - On exploite le calcul de probabilités pour inférer les valeurs des $\mathcal{X}_i$ en fonction des valeurs des $\mathcal{Y}_j$.
- Illustration sur base d'un problème de diagnostic "jouet"
  - Double pile-ou-face bruité.
  - Suffisamment simple pour faciliter les calculs.
  - Suffisamment riche pour mettre en évidence les idées principales.

Double p.o.f. bruité : modélisation (a)



$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{Z}_1$	$\mathcal{Y}_1$
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	0

- $\mathcal{X}_1, \mathcal{X}_2$ sont des v.a. binaires représentant l'issue du lancer.
- $\mathcal{Y}_1$ l'observation (bruitée) qu'on peut faire : est-ce que $\mathcal{X}_1 = \mathcal{X}_2$?
- $\mathcal{Z}_1$ est une variable non observable valant 1 si $\mathcal{Y}_1$ est erronée.
- $\mathcal{Y}_1$ est donc une fonction des variables $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Z}_1$ (cf. tableau).

Double p.o.f. bruité : modélisation (b)

- $\mathcal{X}_1, \mathcal{X}_2$ sont indépendantes.
- $\mathcal{Z}_1$ est indépendante de $\mathcal{X}_1$ et de $\mathcal{X}_2$ et de toute fonction de $\mathcal{X}_1$ et de $\mathcal{X}_2$.
- Les pièces sont équilibrées : $P_{\mathcal{X}_i}(x_i) = 0.5, \forall x_i \in \{0, 1\}$.
- La probabilité p d'erreur de mesure vaut 0.1 (i.e. $P_{\mathcal{Z}_1}(1) = 0.1$).
- Comment avancer ?
  - Approche théorique :
    trouver un $(\Omega, \mathcal{E}, P)$ compatible avec les hypothèses.
  - Approche pratique :
    modéliser directement la loi conjointe $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}$

NB: nous gardons en tête la nécessité de pouvoir enrichir ultérieurement le modèle en introduisant d'autres variables $\mathcal{Y}_j$ représentant d'autres observations possibles.

Double p.o.f. bruité : modèle théorique $(\Omega, \mathcal{E}, P)$ possible

ω	$P(\{\omega\})$	$\mathcal{X}_1(\omega)$	$\mathcal{X}_2(\omega)$	$\mathcal{Z}_1(\omega)$	$\mathcal{Y}_1(\omega)$
ω_1	0.000	0	0	0	0
ω_2	0.225	0	0	0	1
ω_3	0.025	0	0	1	0
ω_4	0.000	0	0	1	1
ω_5	0.225	0	1	0	0
ω_6	0.000	0	1	0	1
ω_7	0.000	0	1	1	0
ω_8	0.025	0	1	1	1
ω_9	0.225	1	0	0	0
ω_{10}	0.000	1	0	0	1
ω_{11}	0.000	1	0	1	0
ω_{12}	0.025	1	0	1	1
ω_{13}	0.000	1	1	0	0
ω_{14}	0.225	1	1	0	1
ω_{15}	0.025	1	1	1	0
ω_{16}	0.000	1	1	1	1

- Nous avons choisi Ω en postulant un ω par configuration possible des 4 v.a. du problème, et puis nous avons "tiré du chapeau" la loi P.
- On peut se convaincre (cf. développements détaillés dans les notes) que la loi $P(\{\omega\})$ donnée dans la table est bien conforme aux hypothèses du pb.

Modélisation directe de la loi (densité) $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}$

- Premièrement, nous exploitons les informations *structurelles* :

$$\begin{aligned} P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}(x_1,x_2,z_1,y_1) &= P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1}(x_1,x_2,z_1)P_{\mathcal{Y}_1|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1}(y_1|x_1,x_2,z_1) \\ &= P_{\mathcal{X}_1}(x_1)P_{\mathcal{X}_2}(x_2)P_{\mathcal{Z}_1}(z_1)P_{\mathcal{Y}_1|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1}(y_1|x_1,x_2,z_1) \end{aligned}$$

  où nous profitons de l'indépendance mutuelle des v.a. $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Z}_1$, pour remplacer $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1}(x_1,x_2,z_1)$ par $P_{\mathcal{X}_1}(x_1)P_{\mathcal{X}_2}(x_2)P_{\mathcal{Z}_1}(z_1)$.

- Deuxièmement, nous exploitons les informations *quantitatives* :
  - $P_{\mathcal{X}_1}(x_1) = 0.5, \forall x_1 \in \{0,1\}$ (première pièce équilibrée)
  - $P_{\mathcal{X}_2}(x_2) = 0.5, \forall x_2 \in \{0,1\}$ (seconde pièce équilibrée)
  - $P_{\mathcal{Z}_1}(0) = 0.9$ et $P_{\mathcal{Z}_1}(1) = 0.1$ (probabilité d'erreur $p = 0.1$ associée à l'observation $\mathcal{Y}_1$)
  - $P_{\mathcal{Y}_1|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1}(y_1|x_1,x_2,z_1) = 1$ si $y_1 = f_1(x_1,x_2,z_1)$ et $P_{\mathcal{Y}_1|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1}(y_1|x_1,x_2,z_1) = 0$ sinon (f_1, donnée au slide 67).
- Mises ensemble, ces informations structurelles et quantitatives permettent de construire directement la loi conjointe $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}$.

Construction de la densité conjointe $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}$

x_1	x_2	z_1	y_1	$P_{\mathcal{X}_1}$	$P_{\mathcal{X}_2}$	$P_{\mathcal{Z}_1}$	$P_{\mathcal{Y}_1\|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1}$	$P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}$
0	0	0	0	0.5	0.5	0.9	0.0	0.000
0	0	0	1	0.5	0.5	0.9	1.0	0.225
0	0	1	0	0.5	0.5	0.1	1.0	0.025
0	0	1	1	0.5	0.5	0.1	0.0	0.000
0	1	0	0	0.5	0.5	0.9	1.0	0.225
0	1	0	1	0.5	0.5	0.9	0.0	0.000
0	1	1	0	0.5	0.5	0.1	0.0	0.000
0	1	1	1	0.5	0.5	0.1	1.0	0.025
1	0	0	0	0.5	0.5	0.9	1.0	0.225
1	0	0	1	0.5	0.5	0.9	0.0	0.000
1	0	1	0	0.5	0.5	0.1	0.0	0.000
1	0	1	1	0.5	0.5	0.1	1.0	0.025
1	1	0	0	0.5	0.5	0.9	0.0	0.000
1	1	0	1	0.5	0.5	0.9	1.0	0.225
1	1	1	0	0.5	0.5	0.1	1.0	0.025
1	1	1	1	0.5	0.5	0.1	0.0	0.000

$P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}$ est le produit de $P_{\mathcal{X}_1}$, $P_{\mathcal{X}_2}$, $P_{\mathcal{Z}_1}$ et $P_{\mathcal{Y}_1|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1}$

- Cette table fournit une information essentiellement équivalente à celle de la table du slide 69.
- Cependant, elle n'est pas tirée du chapeau : elle est construite de façon systématique en exploitant directement les hypothèses du problème.

Obtention de la densité conjointe $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$

- La variable $\mathcal{Z}_1$, bien qu'utile pour la construction du modèle, ne nous intéresse pas vraiment; en effet, nous sommes seulement intéressé par la densité conjointe $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$.
- $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$ s'obtient en éliminant $\mathcal{Z}_1$ de $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}$, par marginalisation :

$$P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}(x_1, x_2, y_1) = \sum_{z_1 \in \{0,1\}} P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}(x_1, x_2, z_1, y_1).$$

- On calcule $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$ par marginalisation de $\mathcal{Z}_1$ dans $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}_1,\mathcal{Y}_1}$.

x_1	x_2	y_1	$P_{\mathcal{X}_1}$	$P_{\mathcal{X}_2}$	$P_{\mathcal{Y}_1\mid\mathcal{X}_1,\mathcal{X}_2}$	$P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$
0	0	0	0.5	0.5	0.1	0.025
0	0	1	0.5	0.5	0.9	0.225
0	1	0	0.5	0.5	0.9	0.225
0	1	1	0.5	0.5	0.1	0.025
1	0	0	0.5	0.5	0.9	0.225
1	0	1	0.5	0.5	0.1	0.025
1	1	0	0.5	0.5	0.1	0.025
1	1	1	0.5	0.5	0.9	0.225

- $P_{\mathcal{Y}_1\mid\mathcal{X}_1,\mathcal{X}_2}$ s'en déduit en calculant $P_{\mathcal{Y}_1\mid\mathcal{X}_1,\mathcal{X}_2} = P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1} / P_{\mathcal{X}_1,\mathcal{X}_2}$.

Enrichissement : ajout d'une nouvelle observation $\mathcal{Y}_2$

- Nouvelle donnée : content de notre travail, notre chef nous demande d'incorporer dans le modèle une nouvelle v.a. modélisant une nouvelle observation, afin d'évaluer l'intérêt d'un nouveau capteur.
- P.ex. supposons que nous puissions mesurer une v.a. $\mathcal{Y}_2$, qui donne une information (bruitée) sur $\mathcal{X}_1$: disons que $\mathcal{Y}_2 = f_2(\mathcal{X}_1, \mathcal{Z}_2)$ où $\mathcal{Z}_2$ représente l'erreur de mesure du nouveau capteur.
- Modèle de mesure : disons que $\mathcal{Y}_2 = \mathcal{X}_1$ si pas d'erreur de mesure, et que la probabilité d'erreur de mesure est de $p' = 0.2$.
- Disons aussi que $\mathcal{Y}_2$ ne dépend que de $\mathcal{X}_1$, en d'autres termes que l'erreur de mesure $\mathcal{Z}_2$ est indépendante de toutes les autres v.a.
- Pour faire le travail, il nous faut maintenant modéliser la densité conjointe $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1,\mathcal{Y}_2}$. On a

$$P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1,\mathcal{Y}_2} = P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1} P_{\mathcal{Y}_2|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$$

- Le problème se réduit donc à modéliser $P_{\mathcal{Y}_2|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$

Modélisation de la densité $P_{\mathcal{Y}_2|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$

- Montrons d'abord que suite aux hypothèses concernant $\mathcal{Y}_2$, on a

$$P_{\mathcal{Y}_2|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}(y_2|x_1,x_2,y_1) = P_{\mathcal{Y}_2|\mathcal{X}_1}(y_2|x_1). \tag{35}$$

En effet, par hypothèse on doit avoir

$$P_{\mathcal{Y}_2|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}(0|0,x_2,y_1) = 0.8, \tag{36}$$
$$P_{\mathcal{Y}_2|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}(0|1,x_2,y_1) = 0.2, \tag{37}$$
$$P_{\mathcal{Y}_2|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}(1|0,x_2,y_1) = 0.2, \tag{38}$$
$$P_{\mathcal{Y}_2|\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}(1|1,x_2,y_1) = 0.8, \tag{39}$$

quelles que soient les valeurs x_2 de $\mathcal{X}_2$ et y_2 de $\mathcal{Y}_1$ ce qui implique que l'identité (35) est bien vérifiée.

- La densité conjointe $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1,\mathcal{Y}_2}$ peut donc être obtenue directement en multipliant $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$ et $P_{\mathcal{Y}_2|\mathcal{X}_1}$.

Densité conjointe $P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1,\mathcal{Y}_2}$

x_1	x_2	y_1	y_2	$P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1}$	$P_{\mathcal{Y}_2\|\mathcal{X}_1}$	$P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1,\mathcal{Y}_2}$
0	0	0	0	0.025	0.800	0.020
0	0	0	1	0.025	0.200	0.005
0	0	1	0	0.225	0.800	0.180
0	0	1	1	0.225	0.200	0.045
0	1	0	0	0.225	0.800	0.180
0	1	0	1	0.225	0.200	0.045
0	1	1	0	0.025	0.800	0.020
0	1	1	1	0.025	0.200	0.005
1	0	0	0	0.225	0.200	0.045
1	0	0	1	0.225	0.800	0.180
1	0	1	0	0.025	0.200	0.005
1	0	1	1	0.025	0.800	0.020
1	1	0	0	0.025	0.200	0.005
1	1	0	1	0.025	0.800	0.020
1	1	1	0	0.225	0.200	0.045
1	1	1	1	0.225	0.800	0.180

- Cette nouvelle densité permet de faire de nouvelles inférences.
- P.ex., sachant que nous observons $\mathcal{Y}_1 = 1$ (les deux pièces probablement tombées du même côté) et que $\mathcal{Y}_2 = 0$ (la première pièce est probablement tombée sur pile, i.e. $\mathcal{X}_1 = 0$), quelle est la probabilité que la seconde pièce soit tombée sur pile, i.e. quelle est la valeur de $P_{\mathcal{X}_2|\mathcal{Y}_1,\mathcal{Y}_2}(0|1,0)$?
  On calcule à partir de la Table que $P_{\mathcal{X}_2,\mathcal{Y}_1,\mathcal{Y}_2}(0,1,0) = 0.180 + 0.005 = 0.185$, et que
  $P_{\mathcal{Y}_1,\mathcal{Y}_2}(1,0) = 0.180 + 0.020 + 0.005 + 0.045 = 0.250$, et on en déduit que
  $P_{\mathcal{X}_2|\mathcal{Y}_1,\mathcal{Y}_2}(0|1,0) = \frac{0.185}{0.250} = 0.740$.

Synthèse

La construction de modèles probabilistes nécessite de faire évoluer le modèle au fur et à mesure en introduisant de nouvelles variables aléatoires. Une façon systématique de construire un modèle probabiliste est la suivante:

- On introduit progressivement les v.a. dans le modèle, en exploitant les indépendances (connaissances structurelles) entre la nouvelle variable et celles déjà introduites, pour ajouter un nouveau facteur aussi simple que possible à multiplier avec la densité conjointe des variables déjà traitées.
- On introduit, quand il faut faire des inférences probabilistes, les données numériques et les relations fonctionnelles entre variables, pour définir les densités élémentaires qui interviennent comme facteurs dans la densité conjointe concernée par l'inférence probabiliste qu'on souhaite effectuer.
- On exploite ensuite les opérations de marginalisation pour simplifier les densités conjointes afin d'en éliminer des variables, et pour calculer les probabilités associées aux valeurs des variables d'intérêt.
- On exploite la formule de Bayes pour établir des densités conditionnelles pas directement déductibles des hypothèses décrivant le problème, et pour répondre à des questions spécifiques d'inférence probabiliste.

4.4.2 Marginalisation et conditionnement (synthèse)

Au menu

- Dans cette partie, nous partons de la loi (plus précisément, la densité) conjointe sur un certain nombre de v.a. et montrons comment l'exploiter.
- Nous montrons comment, à partir de cette densité, on peut obtenir les densités conjointes d'un sous-ensemble de v.a., et les densités conditionnelles d'un sous-ensemble de v.a. relatives à un autre sous-ensemble de v.a.
- Nous discutons ensuite la notion "d'indépendance conditionnelle".
- Finalement, nous illustrons la notion de modèle probabiliste graphique, qui représente une avancée principale du calcul de probabilité au cours des 30 dernières années pour la structuration de toutes ces idées.

Soit $\{\mathcal{X}_1, \ldots, \mathcal{X}_p\}$ un ensemble de variables aléatoires discrètes définies sur un même espace de probabilité $(\Omega, \mathcal{E}, P)$ et soit $P_{\mathcal{X}_1,\ldots,\mathcal{X}_p}(x_1, \ldots, x_p)$ la densité de probabilité conjointe de ces variables définie pour toute configuration $x_1, \ldots, x_p$ des valeurs des variables $\mathcal{X}_1, \ldots, \mathcal{X}_p$.

Marginalisation

L'opération de marginalisation consiste à éliminer une variable, disons $\mathcal{X}_k$ de $\{\mathcal{X}_1, \ldots, \mathcal{X}_p\}$ afin de déduire la densité conjointe des autres variables. Nous noterons $\{\mathcal{X}_1, \ldots, [\mathcal{X}_k], \ldots, \mathcal{X}_p\}$ l'ensemble des autres variables.

La densité conjointe des variables $\{\mathcal{X}_1, \ldots, [\mathcal{X}_k], \ldots, \mathcal{X}_p\}$ s'obtient par le théorème des probabilités totales:

$$P_{\mathcal{X}_1,\ldots,[\mathcal{X}_k],\ldots,\mathcal{X}_p}(x_1, \ldots, [x_k], \ldots, x_p) = \sum_{x_k \in \Omega_{\mathcal{X}_k}} P_{\mathcal{X}_1,\ldots,\mathcal{X}_k,\ldots,\mathcal{X}_p}(x_1, \ldots, x_k, \ldots, x_p), \tag{40}$$

où la somme porte sur toutes les valeurs possibles de la variable $\mathcal{X}_k$.

Partant de $P_{\mathcal{X}_1,\ldots,\mathcal{X}_p}(x_1, \ldots, x_p)$, on peut donc marginaliser successivement plusieurs variables, pour déduire la densité conjointe d'un sous-ensemble quelconque de $\{\mathcal{X}_1, \ldots, \mathcal{X}_p\}$.

L'ordre dans lequel on élimine les variables n'a pas d'importance étant donnée la commutativité de l'opération de sommation.

Conditionnement simple

La densité conditionnelle des variables $\{\mathcal{X}_1, \ldots, [\mathcal{X}_k], \ldots, \mathcal{X}_p\}$ étant donnée la variable $\mathcal{X}_k$ est définie par

$$P_{\mathcal{X}_1,\ldots,[\mathcal{X}_k],\ldots,\mathcal{X}_p|\mathcal{X}_k}(x_1, \ldots, [x_k], \ldots, x_p|x_k) \triangleq \frac{P_{\mathcal{X}_1,\ldots,\mathcal{X}_k,\ldots,\mathcal{X}_p}(x_1, \ldots, x_k, \ldots, x_p)}{P_{\mathcal{X}_k}(x_k)}. \quad (41)$$

$\forall x_k : P_{\mathcal{X}_k}(x_k) > 0$

Son obtention se résume par conséquent à calculer $P_{\mathcal{X}_k}(x_k)$ (par marginalisation des autres variables) et à faire ensuite une division. On peut conditionner sur plusieurs variables, soit en une seule opération, soit en effectuant successivement cette opération dans un ordre quelconque.

Cependant, lorsqu'on conditionne une densité conditionnelle sur une nouvelle variable, il faut le faire en divisant par la densité *conditionnelle* de cette variable par rapport aux variables préalablement dans le conditionnement!

Conditionnement multiple

Par exemple :

$$P_{\mathcal{X}_1,\ldots,[\mathcal{X}_{k_1},\mathcal{X}_{k_2}],\ldots\mathcal{X}_p|\mathcal{X}_{k_1},\mathcal{X}_{k_2}}(x_1,\ldots,[x_{k_1},x_{k_2}],\ldots,x_p|x_{k_1},x_{k_2}) \triangleq \frac{P_{\mathcal{X}_1,\ldots,\mathcal{X}_p}(x_1,\ldots,x_p)}{P_{\mathcal{X}_{k_1},\mathcal{X}_{k_2}}(x_{k_1},x_{k_2})}.$$

Cette densité peut-être obtenue en conditionnnant d'abord sur $\mathcal{X}_{k_1}$ dans $P_{\mathcal{X}_1,\ldots,\mathcal{X}_p}$, puis en conditionnant sur $\mathcal{X}_{k_2}$ dans $P_{\mathcal{X}_1,\ldots,[\mathcal{X}_{k_1}],\ldots,\mathcal{X}_p|\mathcal{X}_{k_1}}$, ou bien dans l'autre ordre. En effet, on a

$$\begin{aligned}\frac{P_{\mathcal{X}_1,\ldots,\mathcal{X}_p}(x_1,\ldots,x_p)}{P_{\mathcal{X}_{k_1},\mathcal{X}_{k_2}}(x_{k_1},x_{k_2})} &= \frac{P_{\mathcal{X}_1,\ldots,[\mathcal{X}_{k_1}],\ldots,\mathcal{X}_p|\mathcal{X}_{k_1}}(x_1,\ldots,[x_{k_1}],\ldots,x_p|x_{k_1})}{P_{\mathcal{X}_{k_2}|\mathcal{X}_{k_1}}(x_{k_2}|x_{k_1})}\\ &= \frac{P_{\mathcal{X}_1,\ldots,[\mathcal{X}_{k_2}],\ldots,\mathcal{X}_p|\mathcal{X}_{k_2}}(x_1,\ldots,[x_{k_2}],\ldots,x_p|x_{k_2})}{P_{\mathcal{X}_{k_1}|\mathcal{X}_{k_2}}(x_{k_1}|x_{k_2})},\end{aligned}$$

puisque
$P_{\mathcal{X}_{k_1},\mathcal{X}_{k_2}}(x_{k_1},x_{k_2}) = P_{\mathcal{X}_{k_2}|\mathcal{X}_{k_1}}(x_{k_2}|x_{k_1})P_{\mathcal{X}_{k_1}}(x_{k_1}) = P_{\mathcal{X}_{k_1}|\mathcal{X}_{k_2}}(x_{k_1}|x_{k_2})P_{\mathcal{X}_{k_2}}(x_{k_2})$.

Indépendance conditionnelle (définition)

On dit que la variable $\mathcal{X}$ est indépendante de la variable $\mathcal{Y}$ conditionnellement à la variable $\mathcal{Z}$, ce que l'on note par $\mathcal{X} \perp \mathcal{Y}|\mathcal{Z}$, si

$$\forall x, y : P_{\mathcal{X},\mathcal{Y}|\mathcal{Z}}(x, y|z) = P_{\mathcal{X}|\mathcal{Z}}(x|z)P_{\mathcal{Y}|\mathcal{Z}}(y|z), \quad (42)$$

pour toute valeur de z telle que $P_{\mathcal{Z}}(z) > 0$.

(Les densités conditionnelles ne sont en effet pas définies lorsque $P_{\mathcal{Z}}(z) = 0$.)

Lorsque $\mathcal{X} \perp \mathcal{Y}|\mathcal{Z}$ on a $P_{\mathcal{X},\mathcal{Y},\mathcal{Z}}(x, y, z) = P_{Z}(z)P_{\mathcal{X}|\mathcal{Z}}(x|z)P_{\mathcal{Y}|\mathcal{Z}}(y|z)$, si $P_{\mathcal{Z}}(z) > 0$ et cette identité reste valable pour un choix arbitraire de $P_{\mathcal{X}|\mathcal{Z}}(x|z)$ et $P_{\mathcal{Y}|\mathcal{Z}}(y|z)$ lorsque $P_{\mathcal{Z}}(z) = 0$. Cela nous permet d'écrire que

$$\mathcal{X} \perp \mathcal{Y}|\mathcal{Z} \Leftrightarrow P_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \overset{p.s.}{=} P_{\mathcal{Z}}P_{\mathcal{X}|\mathcal{Z}}P_{\mathcal{Y}|\mathcal{Z}}, \quad (43)$$

où l'égalité est interprétée de façon fonctionnelle au sens "presque sûrement", et reste valable quels que soient les choix arbitraires concernant les lois conditionnelles pour les valeurs de z telles que $P_{\mathcal{Z}}(z) = 0$.
Notons aussi que la définition (42) implique que si $\mathcal{X} \perp \mathcal{Y}|\mathcal{Z}$ alors

$$\forall x : P_{\mathcal{X}|\mathcal{Y},\mathcal{Z}}(x|y, z) = P_{\mathcal{X}|\mathcal{Z}}(x|z) \quad (44)$$

pour tout couple de valeurs y, z tel que $P_{\mathcal{Y},\mathcal{Z}}(y, z) > 0$, ainsi que

$$\forall y : P_{\mathcal{Y}|\mathcal{X},\mathcal{Z}}(y|x, z) = P_{\mathcal{Y}|\mathcal{Z}}(y|z) \quad (45)$$

pour tout couple de valeurs x, z tel que $P_{\mathcal{X},\mathcal{Z}}(x, z) > 0$.

Il s'en suit que lorsque $\mathcal{X} \perp \mathcal{Y}|\mathcal{Z}$, l'on a aussi $P_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \stackrel{p.s.}{=} P_{\mathcal{Y}} P_{\mathcal{Z}|\mathcal{Y}} P_{\mathcal{X}|\mathcal{Z}}$ et $P_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \stackrel{p.s.}{=} P_{\mathcal{X}} P_{\mathcal{Z}|\mathcal{X}} P_{\mathcal{Y}|\mathcal{Z}}$.

En résumé,

$$
\begin{aligned}
\mathcal{X} \perp \mathcal{Y}|\mathcal{Z} & \Leftrightarrow & \\
\forall x, y, z : P_{\mathcal{X},\mathcal{Y},\mathcal{Z}}(x, y, z) & = & P_{\mathcal{Z}}(z) P_{\mathcal{X}|\mathcal{Z}}(x|z) P_{\mathcal{Y}|\mathcal{Z}}(y|z) & \quad (46) \\
& = & P_{\mathcal{Y}}(y) P_{\mathcal{Z}|\mathcal{Y}}(z|y) P_{\mathcal{X}|\mathcal{Z}}(x|z) & \quad (47) \\
& = & P_{\mathcal{X}}(x) P_{\mathcal{Z}|\mathcal{X}}(z|x) P_{\mathcal{Y}|\mathcal{Z}}(y|z). & \quad (48)
\end{aligned}
$$

L'équation (46) est obtenue en multipliant les deux membres de (42) par $P_{\mathcal{Z}}(z)$, puis le passage à (47) est obtenu en observant que $P_{\mathcal{Z}}(z) P_{\mathcal{Y}|\mathcal{Z}}(y|z) = P_{\mathcal{Y},\mathcal{Z}}(y, z) = P_{\mathcal{Y}}(y) P_{\mathcal{Z}|\mathcal{Y}}(z|y)$, et enfin le passage à (48) se déduit de $P_{\mathcal{Z}}(z) P_{\mathcal{X}|\mathcal{Z}}(x|z) = P_{\mathcal{X},\mathcal{Z}}(x, z) = P_{\mathcal{X}}(x) P_{\mathcal{Z}|\mathcal{X}}(x|y)$.

La notion d'indépendance conditionnelle se généralise à des ensembles de variables aléatoires disjoints (cf syllabus).

Exploitation pour la construction de la densité conjointe

La densité conjointe d'un certain nombre de variables aléatoires, disons $\{\mathcal{X}_i\}_1^p$ peut se **factoriser** de manière générique sous la forme

$$P_{\mathcal{X}_1,\ldots,\mathcal{X}_p} = P_{\mathcal{X}_1} P_{\mathcal{X}_2|\mathcal{X}_1} P_{\mathcal{X}_3|\mathcal{X}_1,\mathcal{X}_2} \cdots P_{\mathcal{X}_p|\mathcal{X}_1,\ldots,\mathcal{X}_{p-1}}, \tag{49}$$

et cette factorisation, dont les facteurs dépendent de l'ordre dans lequel on prend les variables, est valide quel que soit cet ordre.

Pour un ordre fixé de factorisation, on peut exploiter les relations d'indépendances conditionnelles entre les variables du problème, afin de simplifier cette expression. Par exemple, si on peut assurer que $\mathcal{X}_i \perp \mathcal{X}_1|\{\mathcal{X}_i\}_2^{i-1}$ on peut remplacer le facteur $P_{\mathcal{X}_i|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1}}$ par le facteur $P_{\mathcal{X}_i|\mathcal{X}_2,\ldots,\mathcal{X}_{i-1}}$ dans cette formule.
Afin de simplifier au maximum, on peut essayer de déterminer pour chaque facteur $P_{\mathcal{X}_i|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1}}$, un sous-ensemble de taille minimale de variables de l'ensemble $\{\mathcal{X}_i\}_1^{i-1}$, disons $Pa(\mathcal{X}_i, \{\mathcal{X}_1\}_1^{i-1})$ tel que

$$\mathcal{X}_i \perp (\{\mathcal{X}_1\}_1^{i-1} \setminus Pa(\mathcal{X}_i, \{\mathcal{X}_1\}_1^{i-1}))|Pa(\mathcal{X}_i, \{\mathcal{X}_1\}_1^{i-1}).$$

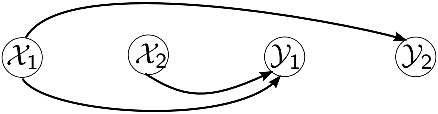
Une fois qu'on a déterminé pour chaque variable $\mathcal{X}_i$ un tel ensemble $Pa(\mathcal{X}_i, \{\mathcal{X}_1\}_1^{i-1})$, on peut représenter la loi jointe sous la forme suivante:

$$P_{\mathcal{X}_1,\ldots,\mathcal{X}_p} = P_{\mathcal{X}_1} P_{\mathcal{X}_2|Pa(\mathcal{X}_2,\{\mathcal{X}_i\}_1^1)} P_{\mathcal{X}_3|Pa(\mathcal{X}_3,\{\mathcal{X}_i\}_1^2)} \cdots P_{\mathcal{X}_p|Pa(\mathcal{X}_p,\{\mathcal{X}_i\}_1^{p-1})}. \tag{50}$$

Indépendances vues d'un point de vue graphique

Deux réseaux bayésiens pour le double pile ou face bruité :

$$P_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}_1,\mathcal{Y}_2} = P_{\mathcal{X}_1}P_{\mathcal{X}_2}P_{\mathcal{Y}_1|\mathcal{X}_1,\mathcal{X}_2}P_{\mathcal{Y}_2|\mathcal{X}_1}$$



$$P_{\mathcal{Y}_1,\mathcal{Y}_2,\mathcal{X}_2,\mathcal{X}_1} = P_{\mathcal{Y}_1}P_{\mathcal{Y}_2|\mathcal{Y}_1}P_{\mathcal{X}_2|\mathcal{Y}_1,\mathcal{Y}_2}P_{\mathcal{X}_1|\mathcal{Y}_1,\mathcal{Y}_2,\mathcal{X}_2}$$

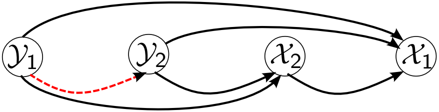


Figure : L'ordre de factorisation a de l'importance !

Chapitre 5 : ensembles de v.a. conjointement gaussiennes

Anticipation sur la dernière leçon

- Construction du modèle : sur base d'un exemple
- Exploitation du modèle : espérance et variances conditionnelles
- Notion de processus stochastique...

Examen écrit :

- Partie 1 (35%):
  - A livres fermés
  - Une question sur le travail 3
  - Trois questions sur la théorie (avec questions subsidiaires courtes relatives à des cas particuliers).
  - Liste de questions de théorie publiée sur la page web du cours avant la fin avril, et commentée lors du dernier cours (14/5/2013).
- Partie 2 (45%):
  - 3 exercices du type de ceux faits lors des séances de répétitions.
  - A livres fermés, mais formulaire sera fourni avec les énoncés.
- Timing :
  - Jeudi 6 juin (500/B7a): début à 8h; fin à 12h30.
  - Pause de 10min entre partie 1 et partie 2.
- Pour mémoire : les notes obtenues pour les travaux pratiques compteront pour 20%