Économétrie II

Ch. 1. Rappels de Concepts

L3 Économétrie – L3 MASS

Pr. Philippe Polomé, U. Lyon 2

Année 2015-2016

Table des matières

Notations

- $Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + \epsilon_i$
  - $i = 1...I$ indexe les observations
    - t séries temporelles, n ou i coupes transversales
  - Y variable endogène, expliquée, dépendante
  - X_j avec $j = 1...k$ variable explicative ou causale, régresseur
    - pas nécessairement exogène
- $E(Y_i) = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki}$
  - (sans le ϵ) **théorie (causale)**
  - β_j mesure quantitativement l'influence de X_j sur Y
    - β_j **pente** de la droite selon X_j
    - β_0 constante, terme indépendant ou intercept
  - β vecteur de coefficients du modèle

Notations

- Terme d'erreur ϵ ou μ
  - Interprétations : erreurs de mesures, régresseurs inobservables ou manquants, facteurs aléatoires...
  - Tous les modèles économétriques sont **stochastiques**
- Notation matricielle $Y = X\beta + \epsilon$
  - On observe Y et X
  - L'économétrie est un ensemble de techniques pour **estimer** β à partir de Y et de X
- Chaque technique = une **formule**, qu'on appelle « **estimateur** » et qu'on note avec un chapeau
  - Exemple « Moindres carrés ordinaire » $\hat{\beta} = \left(X'X\right)^{-1} X'Y$
  - $\exists$ estimateurs « maximum de vraisemblance », « méthode des moments », « variables instrumentales », « moindres carrés généralisés »...

Notations

- **Prédiction** avec un chapeau $\hat{Y}$ $\forall$ les estimateurs
  - $\hat{Y} = X\hat{\beta}$ avec MCO p.e.
    - Toujours erreur de prédiction
  - $X \in$ échantillon : "**Valeur ajustée**"
- On n'observe jamais le terme d'erreur ϵ
  - $(Y, X) \in$ échantillon : calcul du **résidu** $\hat{\epsilon} = Y - \hat{Y} = Y - X\hat{\beta}$
- Résidu s'écrit parfois e et peut se nommer « erreur »
  - Porte à confusion

Exemple simple

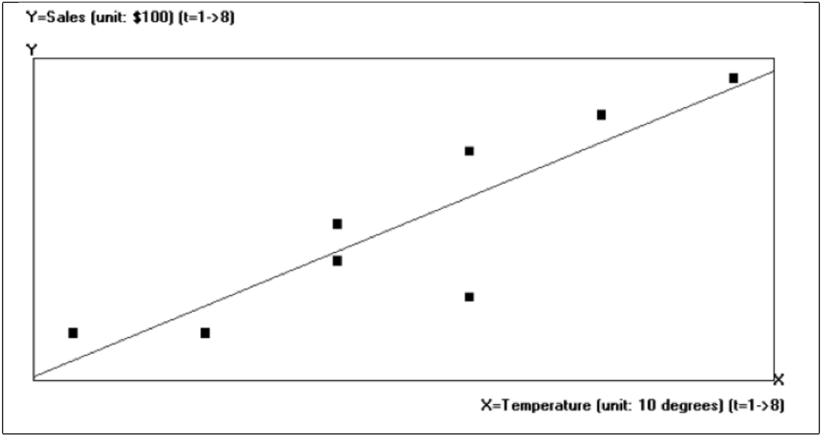
Table 1: *Ice cream data*

Sales (unit = \$100)	Temperature (unit = 10 degrees)
$Y_1 = 8$	$X_1 = 5$
$Y_2 = 10$	$X_2 = 7$
$Y_3 = 8$	$X_3 = 6$
$Y_4 = 13$	$X_4 = 8$
$Y_5 = 15$	$X_5 = 10$
$Y_6 = 14$	$X_6 = 9$
$Y_7 = 11$	$X_7 = 7$
$Y_8 = 9$	$X_8 = 8$

Ce qu'on a et ce qu'on veut faire

- Modèle **causal stochastique**
  - S'il fait plus chaud (X : température), la consommation de crème glacée (Y) augmente, toutes autres choses (ϵ) égales
  - $Y_t = \beta_1 + \beta_2 X_t + \epsilon_t$
    - pour chaque $t = 1 \ldots T$
    - avec erreurs aléatoires non-observables ϵ_t (pas toute autres choses égales)
    - Série temporelle courte $\simeq$ coupe transversale
- On a des données
  - Variable Y
  - Une ou plusieurs variables $X_1 \ldots X_k$ explicatives
- On veut
  - Quantifier l'influence de X sur Y
  - Prédire Y conditionnellement à certaines valeurs pour X
- Comment faire ?

Intuition 1 : Droite qui "passe au mieux"



Intuition 1 : Droite qui "passe au mieux"

- $Y_t = \beta_1 + \beta_2 X_t + \epsilon_t$ pour chaque $t = 1 \dots T$
- On suppose que c'est une bonne approximation de la relation réelle
  - ϵ = événement aléatoire **non-mesuré** et non-systématique
  - ϵ pas corrélé avec X
- $\hat{\beta}$ de β t.q. la droite $Y_t = \hat{\beta}_1 + \hat{\beta}_2 X_t$ "passe au mieux" dans le nuage de points
  - C'est du dessin : minimisation des distances euclidiennes
  - Soit le résidu $\hat{\epsilon} = Y - X\hat{\beta}$
  - Cherche le vecteur de nombres $\hat{\beta}$ t.q. somme des carrés des résidus $\sum_{t=1}^{T} \left(Y_t - X_t\hat{\beta}\right)^2$ est minimale
  - Réponse $\hat{\beta} = \left(X'X\right)^{-1} X'Y$: estimateur **moindres carrés**

Intuition 1 : Logiciels

- L'exemple dans Tableur Icecream_inverse_TCD.ods
  - Dessin
  - Calcul matriciel
- L'exemple dans Gretl

Intuition 2 : "Inversion" de $Y = X\beta + \epsilon$

- Imaginons que $\epsilon = 0$ et que X soit carrée et inversible
  - Alors $\exists\ X^{-1}$ telle que $X^{-1}X = I$
  - $\beta = X^{-1}Y$ s'obtient par inversion
    - système d'équations linéaires
- Mais $\epsilon \neq 0$ et X pas carrée $\rightarrow$ « Généralisation » de l'inverse
  - Prémultiplier par X', on a $X'Y = X'X\beta + X'\epsilon$
  - $X'X$ est carrée et "souvent" inversible
    - À condition d'absence de multicolinéarité
  - **Hypothèse** : $X'\epsilon = 0$ covariance zéro
  - Alors $X'Y = X'X\beta$ et donc $\beta = \left(X'X\right)^{-1} X'Y$
    - exactement : β et non $\hat{\beta}$

Intuition 2 : “Inversion” de $Y = X\beta + \epsilon$

- En général, $X^{'}\epsilon \neq 0$ mais
  - Si on peut supposer que $X^{'}\epsilon \approx 0$ en un sens stochastique
    - au moins lorsque la taille de l'échantillon $\to \infty$
  - Alors on peut écrire $\hat{\beta} = \left(X^{'}X\right)^{-1} X^{'}Y$
    - dans le sens où $\hat{\beta}$ est une approximation de β quand $n \to \infty$
- L'inversion est une intuition d'un autre estimateur
  - **Méthode des moments**

Estimateur Méthode des Moments MM

- Soit A un estimateur de β, alors on peut écrire $Y - XA = \hat{\epsilon}$
- Hypothèse exogénéité $E(\epsilon|X) = 0$
  - $\implies E(X\epsilon) = 0$ (corrélation nulle $\forall$ régresseurs)
- Stratégie MM
  - Cette **hypothèse** sur les moments de la pop. est **imposée** aux moments de l'échantillon : on veut A t.q. $X'\hat{\epsilon} = 0$
  - Donc : $X'(Y - XA) = 0$: CPO des MCO
- Alors $A = (X'X)^{-1}X'Y = \hat{\beta}$:
  - Estimateur MM = estimateur MCO pour le MRL $Y = X\beta + \epsilon$
    - $X'\hat{\epsilon} = X'\left(Y - X\left(X'X\right)^{-1}X'Y\right) = 0$ par construction
  - Mais la méthode des moments s'applique pareillement à des modèles non linéaires $Y = g(X, \beta) + \epsilon$
  - Et le principe est très différent

Prédiction & effet marginal

- Un objectif était de prédire au mieux les ventes de crème glacée en fonction de la température
- La prédiction se trouve sur la droite : $\hat{Y} = X\hat{\beta}$
  - Aussi bien pour MCO que pour MM dans ce cas-ci
- Un 2^{nd} objectif était la quantification de l'influence de X sur Y
  - ${}^{\partial Y}/_{\partial x_1}$ estimé $= \hat{\beta}_1 =$ coefficient estimé de $X_1 =$ pente de la droite
  - $\beta_1 =$ effet (constant) d'un changement marginal de x_1 sur y

Table des matières

Causalité contre corrélation

- Modèle causal : X cause Y, X influence Y, $X \rightarrowtail y$
  - Pas le contraire
- Ce n'est pas la même chose qu'une corrélation
  - X corrélé à $Y \Leftrightarrow Y$ corrélé à X
- Dans l'exemple des ventes de crème glacée, la température cause les ventes
  - Un accroissement de température provoque un accroissement de demande
  - Ce n'est pas parce que les gens mangent plus de glaces que la température va augmenter

Simultanéité

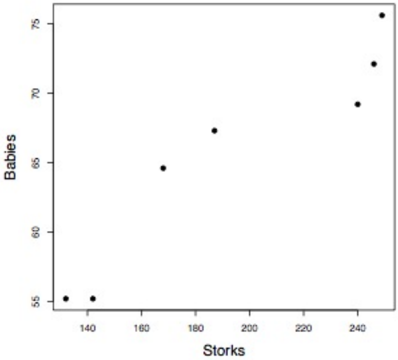
Dans des modèles plus sophistiqués, causalité pas évidente

- En macro, le taux de change agit-il sur la balance commerciale ou est-ce l'inverse ?
- En marketing, au niveau de la firme, les ventes et les dépenses de publicité sont dites simultanées
  - chacune est cause de l'autre
- La demande d'un produit (quantité) dépend du prix, mais l'inverse est vrai aussi (simultanéité)

Régression inverse

- La régression ne mesure **que des corrélations**
  - Ne peut que **confirmer ou infirmer** un modèle théorique
    - Uniquement dans un sens statistique : pour un certain jeu de données, le modèle est confirmé/infirmé
  - $\exists$ tests de causalité applicables dans certaines circonstances
  - Certaines données (panel p.e.) permettent d'être plus sûr de la causalité
    - Changements expliquent changements
- Exemple de régression inverse dans le cas des crèmes glacées : $X_1 = \gamma_1 + \gamma_2 Y + \epsilon$
  - Tableur Icecream_inverse_TCD.ods

Exemple des cigognes



- Fisher, 1936
- Copenhagen, décennie post WWI
- En réalité : constructions importantes et migration des campagnes
- Exemple de régression **fallacieuse** spurieuse (spurious)

Pq les cigognes sont-elles associées aux bébés ?

- Oiseau migrateur européen
- Part à l'automne et revient en Europe centre & nord début avril
  - Soit 9 mois après le solstice d'été (21 juin / Saint Jean) environ
- Le solstice d'été était un important festival païen, où les gens se mariaient beaucoup

Les moyennes conditionnelles

- Exemple de tableau croisé dynamique / Pivot table dans Excel
  - Google “Excel 25 easy PivotTable reports”
  - Tableur Icecream_inverse_TCD.ods
- Les moyennes conditionnelles sont des moyennes (arithmétiques simples) calculées **par groupe** dans l'échantillon
  - P.e. les ventes moyennes par vendeur dans une entreprise
- Il est facile de croire que les différences entre moyennes conditionnelles sont dues aux “conditionneurs”
  - P.e. les ventes de Smith sont plus élevées que celles de González parce que Smith est meilleur vendeur que González
- Mais il ne s'agit que d'une corrélation, pas d'une causalité
  - P.e. les ventes de Smith sont plus élevées parce qu'il est sur un plus grand territoire
  - Pas “toutes autres choses égales” / *ceteris paribus*

Causalité : conclusion

- L'économétrie, les moyennes conditionnelles, les corrélations ne servent qu'à **quantifier**, pas à expliquer
- De telles quantifications peuvent confirmer ou infirmer une théorie dans un sens probabiliste
- Il ne faut pas accepter n'importe quel résultat juste parce qu'il a été obtenu par des méthodes sophistiquées
- L'hypothèse de causalité
  - Est commune à tous les modèles économétriques
  - La tester est l'objectif principal de l'économétrie

Table des matières

Chaque échantillon est aléatoire

- Exemple de la vente de crème glacée sur la plage
  - Un autre vendeur sur autre plage aurait récolté des données différentes
  - La méthodologie présentée auparavant est applicable de même
  - Le modèle est le même ... mais les valeurs des coefficients seront différentes dans les deux cas !
- Echantillon aléatoire $\Rightarrow \hat{\beta}$ aléatoire alors que β ne l'est pas !
  - Quel est le $\hat{\beta}$ correct ? Tous les deux sont corrects
  - Tous les deux sont entachés d'une marge d'erreur par rapport au « vrai » coefficient β
- On va illustrer comment fluctuent les $\hat{\beta}$
  - Distribution d'échantillonnage

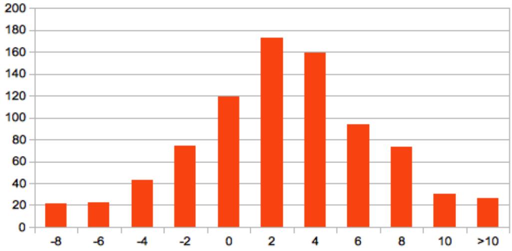
Simulation de Monte-Carlo Monte Carlo.ods

- On génère des données artificielles afin d'illustrer certains outils théoriques dans un cadre contrôlé
- Fonction alea() / rand() : crée une valeur tirée d'une v. a. de distribution uniforme entre 0 et 1
  - sqrt(-2*ln(alea()))*sin(2*pi()*alea()) crée une valeur $n(0,1)$
  - Avec ces fonctions, on génère X et μ
- Calculer $Y = 2 - 3X + \mu$ (ou tout autre choix de coefficient)
- Générer ainsi 10 lignes (par exemple)
- En utilisant Y et X on estime $\hat{\beta}$
  - On voit bien que $\hat{\beta} \neq \beta$
- En recommençant l'opération, on crée des vecteurs $\hat{\beta}_i$ qui sont **tous** différents les uns des autres et de β
  - Les $\hat{\beta}_i$ sont tous aléatoires, β ne l'est pas
  - ϵ est aléatoire
  - X est aléatoire, mais l'analyse est conditionnelle à X
    - Comme si X était constant

Distribution d'échantillonnage

- $\hat{\beta}$ aléatoire : conséquences
  - Pas de garantie d'être proche des vraies valeurs
  - $\hat{\beta}$ suit une **distribution**
    - La valeur de $\hat{\beta}$ change avec chaque échantillon : **distribution d'échantillonnage**
    - Comme toute distribution, elle a des **moments** : moyenne, variance,...
- Selon les valeurs atteintes par ces moments, l'estimateur a des **propriétés** plus ou moins bonnes
  - Un estimateur sera jugé meilleur qu'un autre si ses propriétés sont meilleures
- On va voir pour MCO dans le MRL

Distribution d'échantillonnage de $\hat{\beta}_1$ dans Monte Carlo.ods



- Vraie valeur 1.5
- Moyenne 1.57...
- Écart type 4.55...
- n=835

Modèle de régression linéaire : 7 hypothèses

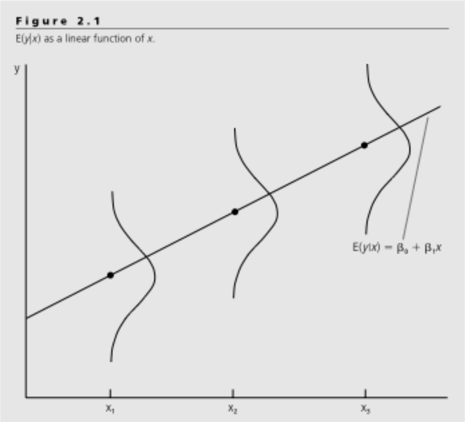
Les circonstances dans lesquelles MCO est un "bon" estimateur

- Modèle de Régression Linéaire (MRL) $Y = X\beta + \epsilon$
  - 7 hypothèses classiques (+ celle de causalité)
- Lorsqu'elles sont vérifiées, l'estimateur MCO possède des propriétés désirables
  - Dans quels cas ne sont-elles pas satisfaites ?
  - Conséquences sur l'estimateur ?
  - Peut-on "réparer" ?
    - Proposer un estimateur alternatif, transformer les données...

MRL $Y = X\beta + \epsilon$: hypothèses 1-4

1. $E(\epsilon_i) = 0 \ \forall i$: les erreurs ont une **espérance nulle**
2. $var(\epsilon_i) = \sigma^2 \ \forall i$: la variance de chaque erreur est la même et est réelle = **Homoscédasticité**
3. $cov(\epsilon_t, \epsilon_s) = 0 \ \forall t \neq s$: les erreurs sont indépendantes entre elles = **Pas d'auto-corrélation**
   - 1+2+3 = "Sphéricité des erreurs"
4. $E(\epsilon_i x_i) = 0 \ \forall i$: il n'y a pas de corrélation contemporaine (même i) entre l'erreur et chaque régresseur = **Exogénéité**

Figure: MRL $Y = X\beta + \epsilon$: Illustration graphique des 4 hyp. sur l'erreur



Tiré de Wooldridge

MRL $Y = X\beta + \epsilon$: hypothèses 5-7

- 5. X de plein rang
  - Aucun régresseur ne peut s'écrire comme une combinaison linéaire des autres régresseurs
  - Sinon : **colinéarité** parfaite des régresseurs
    - $X^{'}X$ pas inversible
- 6. MRL correctement spécifié
  - La réalité est effectivement linéaire en les coefficients β (forme fonctionnelle)
  - β **non stochastiques**
  - Il ne manque aucun régresseur **pertinent**
- 7. Y continue
  - Pas qualitative : 0/1 ou bien A, B, C, D
  - Pas discrète : 0,1,2,3...
  - Pas tronquée/censurée : [3,12] ou [-1,+1]
- Hypothèses MRL pas respectées $\Rightarrow$ MCO perd certaines/toutes propriétés

MRL $Y = X\beta + \epsilon$: Propriétés de l'estimateur MCO

Table: Propriétés de l'estimateur MCO lorsque toutes les hypothèses du MRL sont respectées

Moment*	n petit	$n \to \infty$
Espérance	biais	consistance
Variance	efficience	efficience asymptotique

* de la distribution d'échantillonnage

Propriété 1 de MCO dans le MRL : absence de biais

- L'espérance de l'estimateur $E\left(\hat{\beta}\right) = \beta$
  - L'estimateur est dit **non-biaisé**
  - Preuve math en annexe
- **La moyenne des coefficients estimés (sur l'ensemble des échantillons simulés) tend à se rapprocher des vraies valeurs**
  - « En moyenne, cet estimateur ne se trompe pas »
  - $E\left(\hat{\beta}\right) \approx moyenne\left(\hat{\beta}\right)$ lorsqu'il y a beaucoup d'échantillons
  - Illustré dans Monte Carlo.ods
- $\hat{\beta}_{MCO}$ est non-biaisé

Propriété 2 de MCO dans MRL : consistance / convergence

- **Plus la taille de l'échantillon grandit, plus les coefficients estimés tendent à se rapprocher des vraies valeurs**
  - L'estimateur est dit **consistant**
    - Souvent "**convergent**" en français
  - Lorsque la taille de l'échantillon tend vers l'infini, les coefficients estimés convergent (en probabilité) vers les vraies valeurs
  - On écrit : $Plim\left(\hat{\beta}\right) = \beta$
- Preuve math assez compliquée
  - Plus facile à illustrer dans un Tableur Monte Carlo.ods
- $\hat{\beta}_{MCO}$ est consistant

Propriétés 3-4 de MCO dans le MRL : efficience

- Théorème de **Gauss-Markov** : $var\left(\hat{\beta}\right)$ est la plus petite de tous les estimateurs linéaires non-biaisés
  - $\hat{\beta}_{MCO}$ est **BLUE** = **efficient**
- Théorème de **Cramer-Rao** : $\hat{\beta}$ est le plus efficient de tous les estimateurs consistants
  - $\hat{\beta}_{MCO}$ est **asymptotiquement efficient** = atteint la borne inférieure de Cramer-Rao
- L'efficience d'un estimateur est sa précision
  - Inverse de sa variance
- L'efficience est **comparative**
  - Un estimateur $\hat{\beta}$ est plus efficient qu'un estimateur $\tilde{\beta}$ si $var\left(\tilde{\beta}\right) - var\left(\hat{\beta}\right)$ est une matrice sdp

Devoir #1 : Monte-Carlo

- Réaliser votre propre exemple de Monte Carlo dans un tableur
  - Avec 2 régresseurs, un distribué uniformément dans $[0, 1]$ et l'autre distribué normalement $n(0, 1)$, une constante et un terme d'erreur distribué $n(0, 1)$
  - Choisissez les valeurs des coefficients
- Tout le monde devrait avoir des chiffres différents
- Calculez les coefficients explicitement avec les formules
  - $X'X$ sera 3x3 et $\hat{\beta}$ sera 3x1
- Calculez le R^2
- Répliquer l'opération avec des tailles d'échantillons croissantes pour monter la consistance de l'estimateur
  - Les devoirs ne sont ni notés ni corrigés, mais ils sont matières d'examen
  - Si vous avez des difficultés à les faire, on en discute en cours

Annexe. Démonstration de $E\left(\hat{\beta}\right) = \beta$

$$E\left(\hat{\beta}\right) = E\left(\left(X'X\right)^{-1} X'Y\right)$$

$$= E\left(\left(X'X\right)^{-1} X'\left(X\beta + \epsilon\right)\right)$$

$$= E\left(\left(X'X\right)^{-1} X'X\beta\right) + E_{X,\epsilon}\left(\left(X'X\right)^{-1} X'\epsilon\right)$$

- parce que $E\left(somme\right) = somme\left(E\right)$ et X est une v.a.

$$= E\left(\beta\right) + E_X\left(E_\epsilon\left[\left(X'X\right)^{-1} X'\epsilon | X\right]\right)$$

- par la loi d'itération des espérances (ci-dessous)

$$= E\left(\beta\right) + E_X\left(\left(X'X\right)^{-1} X' E_\epsilon\left[\epsilon | X\right]\right)$$

$$= \beta \text{ si } E_\epsilon\left[\epsilon | X\right] = 0$$

Loi d'itération des espérances

$$E[Y] = E_X\left[E_{Y|X}(Y|X)\right]$$

- $E[Y]$ est l'espérance inconditionnelle (ou marginale) de Y
- $E_X[\,]$ est l'espérance inconditionnelle (marginale) par rapport à X (on traite Y comme fixe)
- $E_{Y|X}(\,)$ est l'espérance conditionnelle de Y par rapport à X
- Y et X appartiennent au même espace de probabilités (ci-dessous)

Preuve dans le cas discret

$$\begin{aligned}
E_X\left[E_{Y|X}(Y|X)\right] &= \sum_x E_{Y|X}(Y|X)\Pr(X=x) \\
&= \sum_x \left[\sum_y y \Pr(Y=y|X=x)\right]\Pr(X=x) \\
&= \sum_x \sum_y y\left(\Pr(Y=y|X=x)\Pr(X=x)\right) \\
&= \sum_x \sum_y y \Pr(Y=y, X=x) \\
&= E[Y]
\end{aligned}$$

Car $\Pr(Y=y|X=x)\Pr(X=x) = \Pr(Y=y, X=x)$: la prob conjointe (Y,X) = prob conditionnelle (Y|X) • prob marginale (X)

Loi d'itération des espérances

Exemple des deux dés (à six faces équiprobables)

$$E\left(Dé_1\right) = E_{Dé_2}\left[E_{Dé_1|Dé_2}\left(Dé_1|Dé_2\right)\right]$$

$$E\left(Dé_1\right) = \sum_{Dé_1=1}^{6} \Pr\left(Dé_1 = dé_1\right) dé_1 = \sum_i \frac{1}{6}\mathrm{i} = 3.5$$

$$\begin{aligned} E_{Dé_2}\left[E_{Dé_1|Dé_2}\left(Dé_1|Dé_2\right)\right] &= \sum_{Dé_2=1}^{6} E_{Dé_1|Dé_2}\left[Dé_1|Dé_2\right] \Pr\left(Dé_2 = dé_2\right) \\ &= \sum_{Dé_2=1}^{6} E_{Dé_1|Dé_2}\left[Dé_1|Dé_2\right] \frac{1}{6} \\ &= \sum_{Dé_2=1}^{6} \left(\sum_{Dé_1|Dé_2=i}^{6} \Pr\left(Dé_1|Dé_2 = i\right) i\right) \frac{1}{6} \\ &= \sum_{Dé_2=1}^{6} \left(\sum_{1}^{6} \frac{1}{6} i\right) \frac{1}{6} = 3.5 \end{aligned}$$

Espace de probabilités

- On considère une "expérience" aux résultats aléatoires, p.e. un lancer de 2 dés.
- L'ensemble de tous les résultats élémentaires : (1,1), (1,2)...(6,6) constitue l'espace d'échantillonnage Ω
- Les évènements sont des combinaisons des résultats élémentaires, p.e. "somme des 2 dés = 10", "au moins un des dés = 3"...
  - L'ensemble de ces évènements se nomme un σ algèbre et est noté $\mathcal{F}$
- La fonction P de mesure de probabilité associe à chaque évènement une probabilité
- Ces trois composants $(\Omega, \mathcal{F}, P)$ constitue l'espace de probabilité