Eléments du Calcul des Probabilités
Chapitre 3: Notion de Variable Aléatoire

Louis Wehenkel

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

MATH0062-1 : 2BacIng, 2BacInf - 12/3/2012

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

Remarques concernant le syllabus...

- Maintenance d'un erratum
- Toutes remarques souhaitées
- Rappel: chapitre 5 incomplet

3.1 Définition générale de la notion de variable aléatoire

- Du point de vue applicatif
  1. Une variable aléatoire représente une information dont la valeur dépend du résultat de l'expérience aléatoire
  2. La valeur de la variable aléatoire peut être observée dans un contexte donné
  3. On souhaite étudier l'ensemble de ses valeurs possibles et exploiter l'information pour faire des inférences probabilistes
- Du point de vue mathématique (voir plus loin, après exemple)
  1. Une variable aléatoire $\mathcal{X}$ est une fonction mesurable définie sur Ω à valeurs dans un ensemble $\Omega_{\mathcal{X}}$ muni d'une σ-algèbre $\mathcal{E}_{\mathcal{X}}$
  2. A partir de $\mathcal{E}_{\mathcal{X}}$, elle induit une σ-algèbre $\mathcal{E}_{\Omega/\mathcal{X}}$ "filtrée" sur Ω
  3. A partir de P, elle induit une mesure image $P_{\mathcal{X}}$ sur $(\Omega_{\mathcal{X}}, \mathcal{E}_{\mathcal{X}})$.
- Remarques :
  - Dans ce chapitre nous étudions principalement les propriétés individuelles des variables aléatoires.
  - L'étude systématique des propriétés conjointes d'ensembles de variables aléatoires fait l'objet des chapitre 4 et 5 de ce cours.

3.1 Notion générale de variable aléatoire: mesurabilité

Notion générale de variable aléatoire

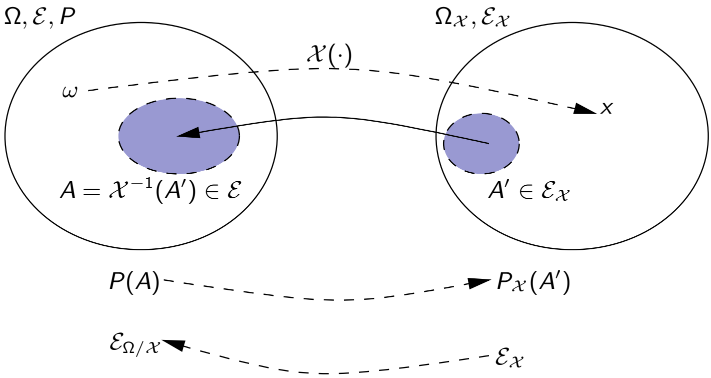
Soient un espace de probabilité $(\Omega, \mathcal{E}, P)$ et une fonction $\mathcal{X}(\cdot)$ définie sur Ω et à valeurs dans un ensemble $\Omega_{\mathcal{X}}$ muni d'une $\sigma-$algèbre $\mathcal{E}_{\mathcal{X}}$. La fonction $\mathcal{X}(\cdot)$ est une variable aléatoire si

$$\forall A' \in \mathcal{E}_{\mathcal{X}} \; : \{\omega \in \Omega | \mathcal{X}(\omega) \in A'\} \in \mathcal{E}. \tag{1}$$

On dit qu'une fonction $\mathcal{X}(\cdot)$ qui vérifie (1) est $(\mathcal{E}, \mathcal{E}_{\mathcal{X}})$-**mesurable** (ou simplement **mesurable**).

Dans la suite nous utiliserons $\mathcal{X}$ pour désigner la variable aléatoire et la notation $\mathcal{X}^{-1}(A')$ pour désigner l'ensemble $\{\omega \in \Omega | \mathcal{X}(\omega) \in A'\}$.

3.1 Notion générale de variable aléatoire



- On a $x \in A' \Leftrightarrow \omega \in A = \mathcal{X}^{-1}(A')$, et $A' \in \mathcal{E}_{\mathcal{X}} \Rightarrow A \in \mathcal{E}$.
- On a $P_{\mathcal{X}}(A') = P(A)$.
- $\mathcal{E}_{\Omega/\mathcal{X}}$ est l'ensemble des parties de Ω s'exprimant comme $\mathcal{X}^{-1}(A')$ avec $A' \in \mathcal{E}_{\mathcal{X}}$

Exemple 2: Lancers de dés

- Expérience 2.1 : Lancer simple

  L'univers : $\Omega = \{1, 2, 3, 4, 5, 6\}$

  Un résultat : $\omega = 3$

- **Expérience 2.2 : Double lancer de dés**

  L'univers : $\Omega = \{(1, 1), (1, 2), \ldots, (5, 6), (6, 6)\}$.
  Un résultat : $\omega = (2, 3)$
  La σ-algèbre : $\mathcal{E} = 2^{\Omega}$
  La mesure de probabilité : $P(\{\omega\}) = 1/36, \forall \omega \in \Omega$

- Expérience 2.3 : Lancer jusqu'au double six

  Univers : $\Omega = \{(6, 6), (1, 6, 6), (2, 6, 6), \ldots\}$,

  Un résultat : $\omega = (1, 3, 2, 4, 4, 5, 1, 2, 2, 2, 5, 4, 6, 6)$

Double lancer de dés : qq variables aléatoires

- Soit $\mathcal{X}(\omega)$ une fonction définie sur Ω dont la valeur est la somme des deux faces supérieures. P.ex. $\mathcal{X}((2,3)) = 5$.
  On a $\mathcal{X}(\Omega) \triangleq \Omega_{\mathcal{X}} = \{2,3,4,5,6,7,8,9,10,11,12\}$
- Soit $\mathcal{Y}_1(\omega)$ une fonction définie sur Ω dont la valeur est la valeur de la face supérieure au premier lancer. P.ex. $\mathcal{Y}_1((2,3)) = 2$.
  On a $\mathcal{Y}_1(\Omega) \triangleq \Omega_{\mathcal{Y}_1} = \{1,2,3,4,5,6\}$
- Soit $\mathcal{Y}_2(\omega)$ une fonction définie sur Ω dont la valeur est la valeur de la face supérieure au second lancer. P.ex. $\mathcal{Y}_2((2,3)) = 3$.
  On a $\mathcal{Y}_2(\Omega) \triangleq \Omega_{\mathcal{Y}_2} = \{1,2,3,4,5,6\}$
- Soit $\mathcal{Z}(\omega)$ une fonction définie sur Ω dont la valeur est 1 si les deux dés portent le même nombre et 0, sinon. P.ex. $\mathcal{Z}((2,3)) = 0$.
  On a $\mathcal{Z}(\Omega) \triangleq \Omega_{\mathcal{Z}} = \{0,1\}$
- Remarque : on a $\forall \omega \in \Omega : \mathcal{X}(\omega) = \mathcal{Y}_1(\omega) + \mathcal{Y}_2(\omega)$
  De même, $\forall \omega \in \Omega : \mathcal{Z}(\omega) = 1 - \min\{1, |\mathcal{Y}_1(\omega) - \mathcal{Y}_2(\omega)|\}$.

Exemple 2: σ-algèbres et mesures de probabilité induites

Pour la suite nous supposons que la mesure P définie sur Ω est uniforme, c'est-à-dire que $P(\{\omega\}) = 1/36, \forall \omega \in \Omega = \{(1,1),(1,2),\ldots,(5,6),(6,6)\}$.

- Etude de la variable $\mathcal{Z}$ (la plus simple...)
  - $\Omega_{\mathcal{Z}} = \{0,1\}$ et $\mathcal{E}_{\mathcal{Z}} = 2^{\Omega_{\mathcal{Z}}} = \{\emptyset, \{0\}, \{1\}, \Omega_{\mathcal{Z}}\}$
  - $\mathcal{E}_{\Omega/\mathcal{Z}} = \{\emptyset, \{(1,2),(2,1),\ldots,(5,6),(6,5)\}, \{(1,1),\ldots(6,6)\}, \Omega\}$
  - On a $P_{\mathcal{Z}}(\{1\}) = 1/6$ et $P_{\mathcal{Z}}(\{0\}) = 5/6$.
- Etude des variables $\mathcal{Y}_1$ et $\mathcal{Y}_2$ (quid de $\mathcal{E}_{\mathcal{Y}_i}?, \mathcal{E}_{\Omega/\mathcal{Y}_i}?$)
  - On a $P_{\mathcal{Y}_1}(\{y_1\}) = 1/6, \forall y_1 \in \Omega_{\mathcal{Y}_1} = \{1,2,3,4,5,6\}$
  - On a $P_{\mathcal{Y}_2}(\{y_2\}) = 1/6, \forall y_2 \in \Omega_{\mathcal{Y}_2} = \{1,2,3,4,5,6\}$
- Etude de la variable $\mathcal{X}$ (la moins simple...)
  - On a $P_{\mathcal{X}}(X) = P(\{\omega \in \Omega : \mathcal{X}(\omega) \in X\}), \forall X \in \mathcal{E}_{\mathcal{X}} = 2^{\{2,3,\ldots,11,12\}}$
  - $\mathcal{E}_{\Omega/\mathcal{X}} = \{\mathcal{X}^{-1}(X) : X \in \mathcal{E}_{\mathcal{X}}\}$.
  - Quelle est la valeur de $P_{\mathcal{X}}(\{6,7,8\})$?
  - Donner un élément de $\mathcal{E}$ qui ne fait pas partie de $\mathcal{E}_{\Omega/\mathcal{X}}$.

3.2 Types de variables aléatoires et de mesures induites

- Variables aléatoires discrètes à valeurs quelconques
  - σ-algèbre : $\mathcal{E}_{\mathcal{X}} = 2^{\Omega_{\mathcal{X}}}$
  - Densité de proba : $p_{\mathcal{X}}(x) \triangleq P(\mathcal{X}(\omega) = x)$.
  - Mesure de proba : $P_{\mathcal{X}}(X) \triangleq P(\mathcal{X}(\omega) \in X) = \sum_{x \in X} p_{\mathcal{X}}(x)$.
  - *Abus de notation :* $P_{\mathcal{X}}(x) \triangleq P_{\mathcal{X}}(\{x\})$
- Variables aléatoires quelconques à valeurs réelles
  - σ-algèbre : les boréliens $\mathcal{B}_{\mathbb{R}}$
  - Fonction de répartition : $F_{\mathcal{X}}(x) \triangleq P(\mathcal{X}(\omega) < x)$.
- Variable aléatoire réelles continues
  - La v.a. est continue si elle admet une densité $f_{\mathcal{X}}(x)$ avec

$$F_{\mathcal{X}}(b) - F_{\mathcal{X}}(a) = \int_a^b f_{\mathcal{X}}(x)\,\mathrm{d}x.$$

  - $F_{\mathcal{X}}$ est alors dérivable et continue et on a $f_{\mathcal{X}}(x) = \frac{\partial F_{\mathcal{X}}(x)}{\partial x}$

Explication : v.a. réelles et σ-algèbre des boréliens $\mathcal{B}_{\mathbb{R}}$

Notion de σ-algèbre borélienne $\mathcal{B}_{\mathbb{R}}$

Par définition, $\mathcal{B}_{\mathbb{R}}$ est la plus petite σ-algèbre qu'on peut définir sur $\mathbb{R}$ et qui contienne l'ensemble de tous les intervalles.

Remarques:

- *$\mathcal{B}_{\mathbb{R}}$ est strictement plus petit que $2^{\mathbb{R}}$, mais suffisamment grand pour tous les besoins pratiques.*
- *Ce choix évite certains paradoxes dont la discussion sort du cadre de ce cours.*
- *On définit de même $\mathcal{B}_{\mathbb{R}^n}$ (cf chapitres 4 et 5).*

"Quels que soient les progrès des connaissances humaines, il y aura toujours place pour l'ignorance et par suite pour le hasard et la probabilité."

Emile Borel, 1871 - 1956

Exemple de fonction de répartition discrète

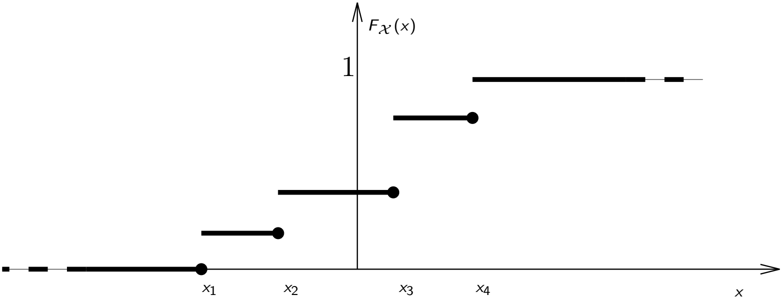


Figure: Exemple de fonction de répartition d'une v.a. réelle discrète. Les discontinuités correspondent aux valeurs possibles x_i de la variable aléatoire (ici au nombre de 4).

Exemple 3: variable aléatoire discrète

Consommation électrique d'un bâtiment (premier modèle)

Physiquement elle est contrainte par le dimensionnement de l'installation électrique et la puissance totale p_{m} de tous les équipements installés dans l'immeuble, mais elle varie de façon aléatoire d'un moment à l'autre.

Supposons que nous disposions d'un enregistrement au cours de l'année précédente, et choisissons au hasard un un instant $t \in [t_{\mathrm{min}}, t_{\mathrm{max}}] \subset \mathbb{R}$ puis observons la consommation à cet instant t. Cette expérience définit une variable aléatoire $\mathcal{X}$, avec $\Omega_{\mathcal{X}} \subset [0, p_{\mathrm{m}}] \subset \mathbb{R}$.

La valeur de la fonction $F_{\mathcal{X}}(x)$ est égale à la proportion du temps où la consommation électrique est inférieure à x. On a $\forall x \leq 0 : F_{\mathcal{X}}(x) = 0$ et $\forall x > p_{\mathrm{m}} : F_{\mathcal{X}}(x) = 1$.

Si nous supposons que les appareils électriques du bâtiment consomment tous une puissance fixe, une fois qu'ils sont branchés, alors la variable consommation totale instantanée sera une variable discrète, les paliers étant définis par les appareils qui sont branchés à un moment particulier.

Par exemple, si nous avons trois appareils de puissance respective $p_1 = 100$, $p_2 = 150$ et $p_3 = 250$ les valeurs possibles pour $\mathcal{X}$ seraient respectivement de $0, 100, 150, 250, 350, 400$ et 500.

Exemple de fonction de répartition continue

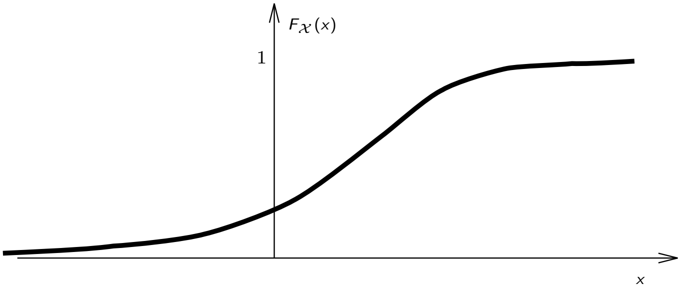


Figure: Exemple de fonction de répartition d'une v.a. réelle continue

Exemple 3: variable aléatoire continue

Consommation électrique d'un bâtiment (modèle plus réaliste)

En pratique, la puissance consommée par un appareil électrique varie continûment :

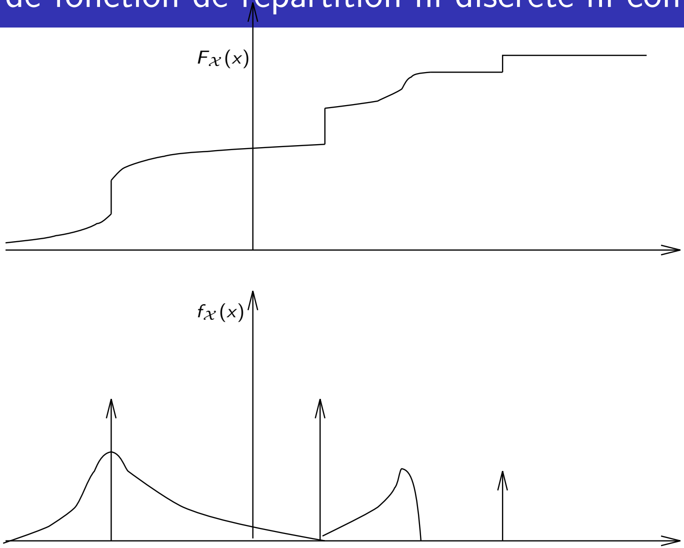
- celle d'un moteur varie en fonction de sa charge, de son accélération;
- celle d'une ampoule électrique ou d'une résistance de chauffage varie en fonction de la tension du réseau électrique (et de la fréquence), l
- a puissance électrique consommée par un ordinateur varie en fonction de la nature des tâches qu'il est en train d'effectuer...

Par ailleurs, le nombre d'équipements d'un bâtiment tel que l'Institut Montefiore est très grand, ce qui veut dire que les paliers sont de toutes façons de très faible amplitude par rapport à la puissance maximale.

Il donc plus réaliste de considérer que la puissance consommée par le bâtiment varie continument au cours du temps, et donc de modéliser $\mathcal{X}$ sous la forme d'une variable aléatoire continue, toutes les valeurs entre O et p_{m} étant en principe possibles.

$F_{\mathcal{X}}(x)$ peut toujours en principe être déterminée pour tout x en vérifiant la proportion du temps où $\mathcal{X} < x$.

Exemple de fonction de répartition ni discrète ni continue



Exemple 3: variable aléatoire mixte

Consommation électrique d'un bâtiment (modèle final)

Pour rendre notre modèle encore plus réaliste, il faut tenir compte des interruptions de fourniture d'électricité et des pannes internes qui arrivent de temps en temps et conduisent à une coupure momentanée de l'alimentation électrique du bâtiment, coupure qui peut avoir une durée variable.

En termes de consommation totale, cela se traduit par une proportion du temps, non nulle, où la consommation totale est nulle, i.e. par une probabilité $P_{\mathcal{X}}(0)$ strictement supérieure à zéro.

La variable $\mathcal{X}$ est dans ce cas ni discrète ni continue.

Sa densité sera composée d'une somme de deux termes, le premier étant une impulsion de Dirac à l'origine de hauteur $P_{\mathcal{X}}(0)$ et et le second étant une densité qui représente la loi associée aux fonctionnements normaux (d'intégrale égale à $1 - P_{\mathcal{X}}(0)$).

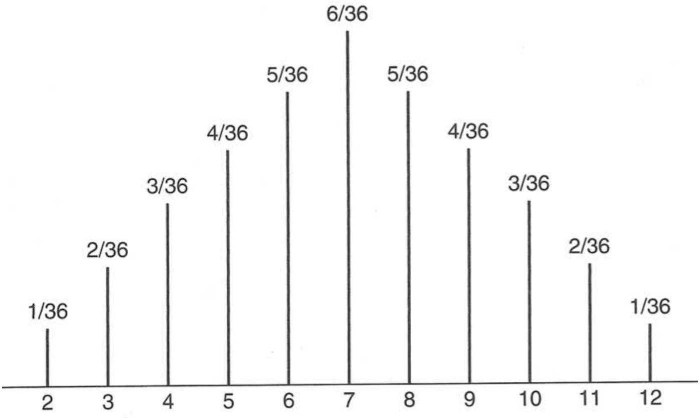
$F_{\mathcal{X}}(x)$ peut cependant toujours en principe être déterminée pour tout x en vérifiant la proportion du temps où $\mathcal{X} < x$. Elle aura une discontinuité à l'origine.

Exemple 2: Calcul des fonctions de répartition

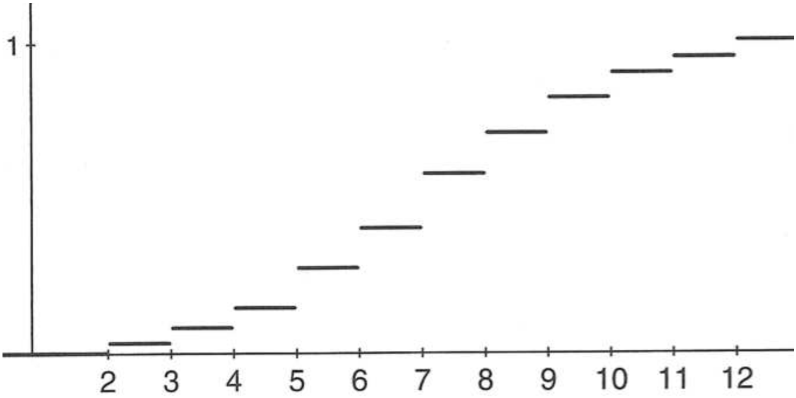
Pour rappel, $P(\{\omega\}) = 1/36, \forall \omega \in \Omega = \{(1,1),(1,2),\ldots,(5,6),(6,6)\}$.

- Variable $\mathcal{Z}$ (la plus simple...)
  - $F_{\mathcal{Z}}(z) = 0, \forall z \leq 0$
  - $F_{\mathcal{Z}}(z) = 5/6, \forall z \in]0,1]$
  - $F_{\mathcal{Z}}(z) = 1, \forall z > 1$.
- Variables $\mathcal{Y}_1$ et $\mathcal{Y}_2$
  - $F_{\mathcal{Y}_1}(y_1) = 0; \forall y_1 \leq 1$
  - $F_{\mathcal{Y}_1}(y_1) = i/6; \forall y_1 \in]i, i+1], i = 1,\ldots 5$
  - $F_{\mathcal{Y}_1}(y_1) = 1; \forall y_1 > 6$.
  - On a (évidemment) $F_{\mathcal{Y}_1}(y) = F_{\mathcal{Y}_2}(y), \forall y \in \mathbb{R}$
  - *Mais cependant, on n'a pas* $\mathcal{Y}_1(\omega) = \mathcal{Y}_2(\omega), \forall \omega \in \Omega$.
  - **Ne pas confondre l'égalité de fonctions de répartition (ou de densités) avec la notion d'égalité de variables aléatoires !!!!!**
- Variable $\mathcal{X}$
  - Faire l'exercice à la maison...

Exemple 2: densité $P_{\mathcal{X}}$ de $\mathcal{X}$



Exemple 2: fonction de répartition $F_{\mathcal{X}}$ de $\mathcal{X}$



3.3 Fonction d'une variable aléatoire

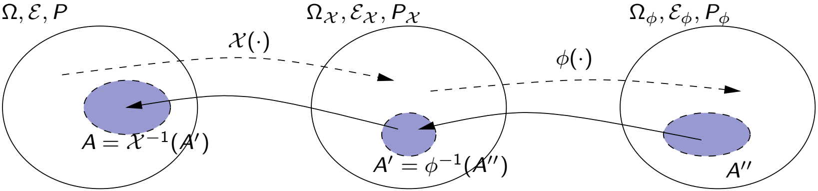


Figure: Toute fonction mesurable d'une v.a. définit une nouvelle v.a.

$A = \mathcal{X}^{-1}(A')$ et $A' = \phi^{-1}(A'')$ $\qquad$ $P_{\phi}(A'') = P_{\mathcal{X}}(A') = P(A)$

Fonction de répartition et densité d'une fonction d'un v.a.

Densité d'une fonction bijective d'une v.a. réelle continue

$$\mathcal{Y} = \phi(\mathcal{X}) \text{ avec } \phi \text{ bijective et dérivable} \Rightarrow f_{\mathcal{Y}}(y) = \frac{f_{\mathcal{X}}(\phi^{-1}(y))}{|\phi'(\phi^{-1}(y))|}. \quad (2)$$

Exemples importants

- $\mathcal{Y} = F_{\mathcal{X}}(x)$ ($\phi = F_{\mathcal{X}}$, et $\phi' = f_{\mathcal{X}}$).
  On obtient que $f_{\mathcal{Y}}(y) = 1$, et on en déduit que la variable $\mathcal{Y}$ possède une densité constante sur l'intervalle $[0, 1]$; il s'agit d'une variable uniforme.
- Réciproquement, si $\mathcal{X}$ est une variable de densité uniforme sur $[0, 1]$, la fonction $\mathcal{Y} = F_{\mathcal{Y}}^{-1}(\mathcal{X})$ possède la fonction de répartition $F_{\mathcal{Y}}$ et la densité $f_{\mathcal{Y}} = F'_{\mathcal{Y}}$ si celle-ci existe.

La dernière formule est utile en pratique, pour générer dans les simulations informatiques des variables aléatoires de distribution donnée, à partir d'un générateur de nombres aléatoires uniformes sur l'intervalle $[0, 1]$.

3.4 Indépendance de deux variables aléatoires

Intuitivement

Deux variables aléatoires $\mathcal{X}$ et $\mathcal{Y}$ sont indépendantes relativement à la mesure P, si tous les événements que l'une induit sur Ω sont indépendants de tous les événements que l'autre induit sur Ω : notion symétrique par définition !

Définition de $\mathcal{X} \perp \mathcal{Y}$ $\quad(\equiv \mathcal{Y} \perp \mathcal{X})$

Deux variables aléatoires $\mathcal{X}$ et $\mathcal{Y}$ définies sur un même espace de probabilité $(\Omega, \mathcal{E}, P)$ sont indépendantes si et seulement si, $\forall A' \in \mathcal{E}_{\mathcal{X}}, \forall B' \in \mathcal{E}_{\mathcal{Y}}$ on a

$$P(\mathcal{X}^{-1}(A') \cap \mathcal{Y}^{-1}(B')) = P(\mathcal{X}^{-1}(A'))P(\mathcal{Y}^{-1}(B')). \quad (3)$$

Nous dirons que la loi induite par la v.a. $\mathcal{Z}(\cdot) = (\mathcal{X}(\cdot), \mathcal{Y}(\cdot))$ sur l'espace produit $\Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ est factorisable *en un produit des lois marginales de $\mathcal{X}$ et de $\mathcal{Y}$* si et seulement si les variables $\mathcal{X}$ et $\mathcal{Y}$ sont indépendantes. On a en effet dans ce cas $\forall A' \in \mathcal{E}_{\mathcal{X}}, \forall B' \in \mathcal{E}_{\mathcal{Y}}$: $P_{\mathcal{X},\mathcal{Y}}((x,y) \in A' \times B') = P_{\mathcal{X}}(x \in A')P_{\mathcal{Y}}(y \in B')$.

Indépendance vs fonctions de répartition et densités

Indépendance $\equiv$ factorisation de la loi jointe

- V.a. discrètes:
  - Densité conjointe : $P_{\mathcal{X},\mathcal{Y}}(x,y) \triangleq P(\mathcal{X}(\omega) = x \wedge \mathcal{Y}(\omega) = y)$
  - $\mathcal{X} \perp \mathcal{Y} \Leftrightarrow P_{\mathcal{X},\mathcal{Y}}(x,y) = P_{\mathcal{X}}(x)P_{\mathcal{Y}}(y), \forall (x,y) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$
- V.a. réelles:
  - F. de rép. conjointe : $F_{\mathcal{X},\mathcal{Y}}(x,y) \triangleq P(\mathcal{X}(\omega) < x \wedge \mathcal{Y}(\omega) < y)$
  - $\mathcal{X} \perp \mathcal{Y} \Leftrightarrow F_{\mathcal{X},\mathcal{Y}}(x,y) = F_{\mathcal{X}}(x)F_{\mathcal{Y}}(y), \forall (x,y) \in \mathbb{R}^2$
- V.a. réelles continues:
  - Densité conjointe: $f_{\mathcal{X},\mathcal{Y}}(x,y) \triangleq \frac{\partial\partial F_{\mathcal{X},\mathcal{Y}}(x,y)}{\partial x \partial y}$
  - $\mathcal{X} \perp \mathcal{Y} \Leftrightarrow f_{\mathcal{X},\mathcal{Y}}(x,y) = f_{\mathcal{X}}(x)f_{\mathcal{Y}}(y), \forall (x,y) \in \mathbb{R}^2$

Indépendance de fonctions de variables aléatoires

- $\mathcal{X} \perp \mathcal{Y} \Rightarrow \phi_1(\mathcal{X}) \perp \phi_2(\mathcal{Y})$
- Quid de la réciproque ? Quid de $\mathcal{X} \perp \phi_1(\mathcal{X})$?

Exemple 2 : illustration de l'indépendance de v.a.

Pour rappel: $P_\Omega(\omega) = 1/36, \forall \omega \in \Omega = \{(1,1),(1,2),\ldots,(5,6),(6,6)\}$.

- *Etude de $\mathcal{Z}$* (qui indique si les deux dés tombent sur la même face)
  - On a $\mathcal{Z} \perp \mathcal{Y}_1$. En effet $\forall y_1 = 1,\ldots,6$ on a :
    $P_{\mathcal{Z},\mathcal{Y}_1}(0,y_1) = 5/36 = (5/6)(1/6) = P_\mathcal{Z}(0)P_{\mathcal{Y}_1}(y_1)$ et
    $P_{\mathcal{Z},\mathcal{Y}_1}(1,y_1) = 1/36 = (1/6)(1/6) = P_\mathcal{Z}(1)P_{\mathcal{Y}_1}(y_1)$
  - de même, on a évidemment $\mathcal{Z} \perp \mathcal{Y}_2$
- *Etude des variables $\mathcal{Y}_1$ et $\mathcal{Y}_2$* (qui indiquent les faces des deux dés)
  - On a $\forall y_1, y_2 = 1,\ldots,6$: $P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1,y_2) = 1/36 = P_{\mathcal{Y}_1}(y_1)P_{\mathcal{Y}_2}(y_2)$
  - Par conséquent $\mathcal{Y}_1 \perp \mathcal{Y}_2$
- *Etude de la variable $\mathcal{X}$* (qui calcule la somme des faces des deux dés)
  - $\mathcal{X} \not\perp \mathcal{Y}_1$: en effet, on a par exemple $P_{\mathcal{X},\mathcal{Y}_1}(12,1) = 0$ alors que $P_\mathcal{X}(12) = 1/36$ et $P_{\mathcal{Y}_1}(1) = 1/6$... (donc on a aussi $\mathcal{X} \not\perp \mathcal{Y}_2$)
  - On a aussi $\mathcal{X} \not\perp \mathcal{Z}$: en effet lorsque $\mathcal{Z} = 1$, la variable $\mathcal{X}$ ne peut prendre un valeur impaire...

  Pourtant, $\mathcal{X} = f(\mathcal{Y}_1,\mathcal{Y}_2)$ et $\mathcal{Z} \perp \mathcal{Y}_1$ et $\mathcal{Z} \perp \mathcal{Y}_2$?!? : explication viendra au chapitre 4...

Indépendance mutuelle de plusieurs v.a.

Anticipation sur le chapitre 4

- Tout comme pour les événements, on distingue aussi pour les variables aléatoires l'indépendance 2-à-2 de l'indépendance mutuelle.
- Pour un ensemble de variables aléatoires $\mathcal{X}_1, \ldots, \mathcal{X}_n$ on dit qu'elles sont mutuellement indépendantes si la mesure conjointe se factorise en produit de mesures marginales, i.e. si $\forall A_1 \in \mathcal{E}_{\mathcal{X}_1}, \ldots, \forall A_n \in \mathcal{E}_{\mathcal{X}_n}$ on a
  - $$\boxed{P\left(\bigcap_{i=1}^n \mathcal{X}_i^{-1}(A_i)\right) = \prod_{i=1}^n P\left(\mathcal{X}_i^{-1}(A_i)\right).}$$
- L'indépendance mutuelle implique l'indépendance deux à deux, mais la réciproque n'est pas vraie.
- Cette propriété implique la factorisation des densités conjointes de tout sous-ensemble de variables aléatoires parmi les $\mathcal{X}_i$, sous la forme d'un produit des densités marginales de ces variables.
- Nous reviendrons sur ces propriétés au chapitres 4 et 5.

3.5 Espérance mathématique d'une v.a. réelle

- Origine du terme
- Commentaires historiques
- Définitions particulières
- • Définition générale
- Propriétés essentielles

Notion d'espérance : origine du terme

- *Dictionnaire de la langue française :*
  Espérance, nom féminin
  Sens 1 : Attente confiante de la réalisation de quelque chose.
  Synonyme : espoir. Anglais : hope.
  Sens 2 : Objet de cette attente.
- *Calcul des probabilités :*
  Espérance mathématique, nom féminin
  Sens 3 : Valeur moyenne d'une variable aléatoire pondérée par sa densité de probabilité.
  Anglais : expectation, expected value, mean.
- *Statistique :*
  Espérance mathématique, nom féminin
  Sens 4 : Grandeur vers laquelle tend, la plupart du temps, la moyenne observée dans un échantillon de grande taille.

NB: dans la suite de ce chapitre nous précisons d'abord le sens 3 puis le sens 4.

Espérance mathématique : commentaires historiques

(source : http://www.math93.com/theoreme/probabilites.html)

Lors d'un voyage à Paris, le physicien et mathématicien hollandais, Christiaan Huygens, prend connaissance de la correspondance entre Fermat et Pascal. Il étudie ces réflexions et publie un traité sur le sujet en 1657, Tractatus de ratiociniis in aleae ludo (Traité sur les raisonnements dans le jeu de dés). C'est le premier traité consacré à cette nouvelle théorie des probabilités.

Huygens comprend déjà, que cette théorie n'est pas uniquement destinée à régler les querelles de joueurs, qu'elle va ouvrir une nouvelle voie et fonder une branche importante des mathématiques modernes.

Le contenu du livre de Huygens, est assez limité mais il y introduit ce qui deviendra la notion d'espérance mathématique.

(source : http://en.wikipedia.org/wiki/Expected_value)

C'est Pierre-Simon Laplace qui bien plus tard, en 1814, dans sa "Théorie analytique des probabilités", définit de manière explicite le concept d'espérance mathématique.

Espérance mathématique : Ω fini

Espérance mathématique d'une v.a. réelle définie sur un espace $(\Omega, \mathcal{E}, P)$ fini

Pour une variable aléatoire réelle définie sur un espace Ω fini, on définit son espérance mathématique (on dit aussi sa moyenne) par

$$E\{\mathcal{X}\} = \mu_{\mathcal{X}} \triangleq \sum_{\omega \in \Omega} \mathcal{X}(\omega) P(\omega) = \sum_{x_k \in \Omega_{\mathcal{X}}} x_k P_{\mathcal{X}}(x_k), \tag{4}$$

où la seconde somme porte sur l'ensemble (fini) des valeurs possibles de $\mathcal{X}$.

Remarque

- Si $1_B(\omega)$ désigne la fonction caractéristique de l'événement B, on a donc $E\{1_B\} = P(B)$.
- Sinon, on peut écrire

$$\mathcal{X}(\omega) = \sum_{x_k \in \Omega_{\mathcal{X}}} x_k 1_{A_k}(\omega), \tag{5}$$

où $A_k = \mathcal{X}^{-1}(x_k)$ désigne le sous-ensemble de Ω où $\mathcal{X}$ vaut x_k et on a

$$E\{\mathcal{X}\} = \sum_{x_k \in \Omega_{\mathcal{X}}} x_k P(A_k) = \sum_{x_k \in \Omega_{\mathcal{X}}} x_k E\{1_{A_k}\}. \tag{6}$$

Exemple 2

- *Etude de $\mathcal{Z}$* (qui indique si les deux dés tombent sur la même face)

$$\begin{aligned} E\{\mathcal{Z}\} &= \mathcal{Z}(1,1)P((1,1)) + \mathcal{Z}(1,2)P((1,2)) + \cdots + \mathcal{Z}(6,6)P((6,6)) \\ &= P((1,1)) + P((2,2)) + \ldots + P((5,5)) + P((6,6)) \\ &= 0P_{\mathcal{Z}}(0) + 1P_{\mathcal{Z}}(1) = 1/6. \end{aligned}$$

- En général, pour une fonction 1_A, où $A \in \mathcal{E}$ on a $E\{1_A\} = P(A)$.
- *Etude des variables $\mathcal{Y}_1$ et $\mathcal{Y}_2$* (qui indiquent les faces des deux dés)
  On a $E\{\mathcal{Y}_1\} = (1+2+3+4+5+6)/6 = 3.5 = E\{\mathcal{Y}_2\}$
- *Etude de la variable $\mathcal{X}$* (qui calcule la somme des faces des deux dés)

$$\begin{aligned} E\{\mathcal{X}\} &= 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + 8\frac{5}{36} + 9\frac{4}{36} \\ &\quad +10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} = 7 \end{aligned}$$

- On remarque que $E\{\mathcal{X}\} = E\{\mathcal{Y}_1\} + E\{\mathcal{Y}_2\}$.

Interprétation : moyenne sur des expériences répétées (a)

- Considérons l'expérience qui consiste à lancer n fois de suite une pièce et supposons que les 2^n résultats possibles soient équiprobables, et donc ont chacun une probabilité de $1/2^n$ de se réaliser.
- Désignons par $\mathcal{X}_i, \forall i = 1, \ldots, n$, une variable aléatoire qui vaut 1 si au i-ème lancer on obtient pile, et 0 sinon.
  - On peut se convaincre que $E\{\mathcal{X}_i\} = 1/2$, $\forall i = 1, \ldots, n$.
    (En effet, sur les 2^n réalisations il y en a 2^{n-1} telles que $\mathcal{X}_i = 1$ et aussi 2^{n-1} telles que $\mathcal{X}_i = 0$.)
  - Les variables $\mathcal{X}_i$ ont donc des lois identiques.
    (On dit qu'elles sont identiquement distribuées : $\forall i = 1, \ldots, n$: $P_{\mathcal{X}_i}(0) = 1/2 = P_{\mathcal{X}_i}(1)$.)
  - On montre aussi que les $\mathcal{X}_i$ sont mutuellement indépendantes.
    (P.ex. on peut voir que $P(\mathcal{X}_k = 1) = P(\mathcal{X}_k = 1 | \mathcal{X}_1 = 1, \ldots, \mathcal{X}_{k-1} = 1)$.)
- Définissons une variable aléatoire $\mathcal{Z}_n$ dont la valeur vaut le nombre relatif de fois k/n que la pièce est tombée sur pile lors d'une expérience. On a
  - $\mathcal{Z}_n = 1/n \sum_{i=1}^{n} \mathcal{X}_i$, et donc $E\{\mathcal{Z}_n\} = 1/n \sum_{i=1}^{n} E\{\mathcal{X}_i\}$
    (étant donnée la linéarité de l'espérance mathématique, cf plus loin...)
  - Donc, $E\{\mathcal{Z}_n\} = 1/2 = E\{\mathcal{X}_i\}$.

Interprétation : moyenne sur des expériences répétées (b)

- Soit $N(n,\epsilon)$ le nombre de réalisations t.q. $\mathcal{Z}_n \in [1/2-\epsilon, 1/2+\epsilon]$. On a évidemment que $N(n,\epsilon) \leq 2^n$, $\forall \epsilon > 0$.
  - Mais on montre que $\lim_{n\to\infty} 2^{-n} N(n,\epsilon) = 1$, $\forall \epsilon > 0$.

    (p.ex. via l'inégalité de Bienaymé-Tchebyshev, voir plus loin...)
- En d'autres termes, lorsque n croit, la moyenne $\mathcal{Z}_n$ des $\mathcal{X}_i$ le long d'une réalisation approche leur espérance commune $E\{\mathcal{X}_i\} = 1/2$.
- Ce résultat constitue la loi faible des grand nombres

  (ici décrite dans le cas des lancers de pièces équilibrées et indépendantes)
- Il existe plusieurs variantes de la loi des grands nombres.
  Ces lois sont à la base de l'interprétation "statistique" de la notion d'espérance mathématique, comme étant la grandeur vers laquelle tend la moyenne observée d'une v.a. lors d'un grand nombre de répétitions indépendantes d'une expérience aléatoire.
- Remarque : notre modèle (loi uniforme sur les 2^n réalisations) implique que les $\mathcal{X}_i$ sont mutuellement indépendantes et identiquement distribuées. Nous disons qu'elles sont i.i.d.

Espérance mathématique : Ω infini et v.a. discrète

Espérance mathématique d'une v.a. réelle discrète quelconque

Pour une variable aléatoire discrète quelconque, on définit son espérance mathématique (on dit aussi sa moyenne) par

$$E\{\mathcal{X}\} = \mu_{\mathcal{X}} \triangleq \sum_{k=1}^{\infty} x_k P_{\mathcal{X}}(x_k), \tag{7}$$

lorsque cette série converge absolument.

NB: lorsque la variable aléatoire prend seulement un nombre fini de valeurs différentes, cette série devient une somme d'un nombre fini de termes.

Espérance mathématique : v.a. continue

Espérance mathématique d'une v.a. réelle continue

Pour une variable aléatoire à valeurs réelles continue on définit son espérance par

$$E\{\mathcal{X}\} = \mu_{\mathcal{X}} = \int_{\mathbb{R}} x f_{\mathcal{X}}(x)\,\mathrm{d}x, \tag{8}$$

lorsque cette intégrale converge absolument.

Cette intégrale n'est pas toujours définie. Par exemple, une variable aléatoire de *Cauchy* (voir plus loin), dont la densité est donnée par

$$f_{\mathcal{X}}(x) = \frac{1}{\pi(1+x^2)}, \tag{9}$$

n'admet pas d'espérance selon la formule (8).

Espérance mathématique : notation générale

L'écriture générale, qu'on rencontre dans de nombreux ouvrages de référence, est la suivante

$$E\{\mathcal{X}\} = \int_{\Omega} \mathcal{X}(\omega)\,\mathrm{d}P(\omega), \tag{10}$$

où le dP indique que l'intégrale est prise par rapport à la mesure P définie sur l'espace de départ Ω, ce qui est équivalent à

$$E\{\mathcal{X}\} = \int_{\mathbb{R}} x\,\mathrm{d}P_{\mathcal{X}}(x), \tag{11}$$

où le $\mathrm{d}P_{\mathcal{X}}$ indique que l'espérance est calculée par rapport à la loi de probabilité induite sur l'espace d'arrivée.

Espérance mathématique : définition générale (a)

V.a. non-négative simple : fonction définie sur Ω qui s'écrit sous la forme

$$\mathcal{Y}(\omega) = \sum_{k=1}^{n} y_k 1_{A_k}(\omega), \tag{12}$$

où les $y_k \in [0, \infty[$ et les A_k sont des événements sur un $(\Omega, \mathcal{E})$ quelconque.

L'espérance mathématique d'une variable aléatoire non-négative simple est

$$E\{\mathcal{Y}\} \triangleq \sum_{k=1}^{n} y_k P(A_k). \tag{13}$$

NB: L'écriture (12) n'est pas unique pour une v.a. simple, mais on peut montrer que la valeur (13) est indépendante de la façon d'exprimer $\mathcal{Y}$ sous la forme (12). Cette définition est évidemment conforme à la définition (4) lorsque Ω est fini, et aussi dans le cas où Ω n'est pas fini mais que la variable aléatoire ne prend qu'un nombre fini de valeurs différentes.

Pour un espace mesurable $(\Omega, \mathcal{E})$ donné, nous désignons par L_{Ω}^{s+} l'ensemble des variables aléatoires non-négatives simples qu'on peut y définir.

Espérance mathématique : définition générale (b)

V.a. non-négative : $\mathcal{X}$ est non-négative si $\forall \omega \in \Omega : \mathcal{X}(\omega) \in [0, \infty]$.
Pour deux v.a. $\mathcal{X}$ et $\mathcal{Y}$ nous écrivons $\mathcal{Y} \leq \mathcal{X}$, si $\mathcal{Y}(\omega) \leq \mathcal{X}(\omega), \forall \omega \in \Omega$.

L'espérance mathématique d'une variable aléatoire non-négative quelconque est

$$E\{\mathcal{X}\} \triangleq \int_{\Omega} \mathcal{X}(\omega) \, \mathrm{d}P(\omega) \triangleq \sup\{E\{\mathcal{Y}\} : \mathcal{Y} \in L_{\Omega}^{s+}, \mathcal{Y} \leq \mathcal{X}\} \leq \infty. \tag{14}$$

Si $E\{\mathcal{X}\} < \infty$ on dit que $\mathcal{X}$ est *P-intégrable*.

NB: Notons que la définition (14) appliquée à une v.a. non-négative simple est équivalente à (13).

Conséquences importantes

- *si* $\mathcal{X} \geq 0$ *alors* $[E\{\mathcal{X}\} = 0] \Leftrightarrow [P(\mathcal{X} = 0) = 1]$, *et* $[E\{\mathcal{X}\} < \infty] \Rightarrow [P(\mathcal{X} = \infty) = 0]$.
- *la somme* $\mathcal{X} + \mathcal{Y}$ *de deux variables aléatoires non-négatives P-intégrables est évidemment encore une variable aléatoire non-négative P-intégrable et* $E\{\mathcal{X} + \mathcal{Y}\} = E\{\mathcal{X}\} + E\{\mathcal{Y}\}$.
- *le produit d'une variable non-négative P-intégrable par une constante* α *positive est encore une variable non-négative P-intégrable et* $E\{\alpha\mathcal{X}\} = \alpha E\{\mathcal{X}\}$.
- *si* $\mathcal{X} \geq \mathcal{Y} \geq 0$ *et que* $\mathcal{X}$ *est intégrable, alors* $\mathcal{Y}$ *l'est aussi, de même que* $\mathcal{X} - \mathcal{Y}$.

Espérance mathématique : définition générale (c)

V.a. réelle quelconque : On a $\mathcal{X} = \mathcal{X}^+ - \mathcal{X}^-$ (non-négatives), avec

$$\mathcal{X}^+(\omega) = \begin{cases} \mathcal{X}(\omega) & \text{si} \quad \mathcal{X}(\omega) \geq 0, \\ 0 & \text{si} \quad \mathcal{X}(\omega) < 0, \end{cases} \quad \text{et} \quad \mathcal{X}^- = (-\mathcal{X})^+. \tag{15}$$

$\mathcal{X}$ est P-intégrable si $E\{\mathcal{X}^+\}$ et $E\{\mathcal{X}^-\}$ sont finies, et son espérance vaut

$$E\{\mathcal{X}\} \triangleq E\{\mathcal{X}^+\} - E\{\mathcal{X}^-\}. \tag{16}$$

Conséquences importantes

- $\mathcal{X}$ est P-intégrable si, et seulement si $|\mathcal{X}| = \mathcal{X}^+ + \mathcal{X}^-$ est P-intégrable. On désigne par L^1_Ω l'ensemble des variables aléatoires réelles P-intégrables pouvant être définies sur $(\Omega, \mathcal{E}, P)$.
- $|E\{\mathcal{X}\}| = |E\{\mathcal{X}^+\} - E\{\mathcal{X}^-\}| \leq |E\{\mathcal{X}^+\}| + |E\{\mathcal{X}^-\}| = E\{|\mathcal{X}|\}$ (et donc $E\{\mathcal{X}\} \leq E\{|\mathcal{X}|\}$)
- Si $\mathcal{X}$ et $\mathcal{Y}$ sont P-intégrables, alors $\mathcal{Z} = \alpha\mathcal{X} + \beta\mathcal{Y}$ est P-intégrable, $\forall \alpha, \beta \in \mathbb{R}$ et on a $E\{\mathcal{Z}\} = \alpha E\{\mathcal{X}\} + \beta E\{\mathcal{Y}\}$. L'ensemble L^1_Ω est donc un espace vectoriel linéaire et l'opérateur d'espérance $E\{\cdot\}$ **est un opérateur linéaire** défini sur cet espace.
- Si $P(\mathcal{X} \neq \mathcal{Y}) = 0$ alors $E\{\mathcal{X}\} = E\{\mathcal{Y}\}$.
- Si $\mathcal{Y}$ peut s'écrire comme une fonction ϕ de $\mathcal{X}$, alors $E\{\mathcal{Y}\} = \int_{\mathbb{R}} \phi(x)\, \mathrm{d}P_{\mathcal{X}}(x)$. C'est le théorème de la mesure image qui justifie l'écriture (11) dans le cas particulier où $\phi(x) = x$.

Homework

NB: la définition générale de la notion d'espérance mathématique que nous venons de donner implique les définitions particulières que nous avons données au préalable.

Cette définition générale est importante pour la cohérence du calcul des probabilités, et c'est de cette définition générale que découlent toutes les propriétés pratiquement importantes que nous allons étudier dans la suite de ce chapitre.

- Lire les pages 3.1 à 3.14 du syllabus.
- Lire très attentivement la section 3.5.2.
- Essayer de se convaincre que les propriétés (conséquences importantes) annoncées dans la section 3.5.2 sont bien vraies, en esquissant le principe d'une démonstration.

Espérance mathématique : propriétés

Au Menu :

- Première version du théorème de l'espérance totale
- Inégalité de Markov
- Espérance d'une fonction d'une v.a.
- ◦ Inégalité de Jensen
- Espérance d'une fonction de plusieurs v.a.

Première version du théorème de l'espérance totale

- Soit un espace $(\Omega, \mathcal{E}, P)$ et une v.a. $\mathcal{X}$ qui est P-intégrable.
- Désignons par $E\{\mathcal{X}|A\}$ l'espérance de la variable $\mathcal{X}$ *sachant que A est réalisé* (i.e. l'espérance $\mathcal{X}$ par rapport à la loi conditionnelle $P(\cdot|A)$).
- Alors, si B est un événement tel que $P(B) >$ et $P(B^c) > 0$, on a :
  - $\mathcal{X}1_B$ et $\mathcal{X}1_{B^c}$ sont aussi P-intégrables (puisque de module majoré par $|\mathcal{X}|$).
  - $E\{\mathcal{X}|B\} = E\{\mathcal{X}1_B\}/P(B)$ (évident si $\mathcal{X}$ est une v.a. non-négative simple).
  - $E\{\mathcal{X}|B^c\} = E\{\mathcal{X}1_{B^c}\}/P(B^c)$ (même remarque).
- De plus :
  - comme $\mathcal{X} = \mathcal{X}1_B + \mathcal{X}1_{B^c}$, $\mathcal{X}$ est donc P-intégrable si et seulement si $\mathcal{X}1_B$ et $\mathcal{X}1_{B^c}$ sont toutes les deux P-intégrables,
  - et on a *(Théorème de l'espérance totale pour une partition binaire)*

$$\boxed{E\{\mathcal{X}\} = P(B)E\{\mathcal{X}|B\} + P(B^c)E\{\mathcal{X}|B^c\}}$$

- Applications : sondages, décomposition de problèmes en sous-problèmes...

Illustration du théorème de l'espérance totale

- *Considérons la v.a.* $\mathcal{Z}$ $(= 1_{(\mathcal{Y}_1=\mathcal{Y}_2)})$ et étudions $E\{\mathcal{Y}_1|\mathcal{Z}\}$

  On a $E\{\mathcal{Y}_1|\mathcal{Z}=1\} = (1+2+3+4+5+6)/6 = 3.5 = E\{\mathcal{Y}_1|\mathcal{Z}=0\}$
  De fait, $\mathcal{Y}_1 \perp \mathcal{Z}$, ce qui implique $E\{\mathcal{Y}_1|\mathcal{Z}\} = E\{\mathcal{Y}_1\}$.

- *Etude de la variable* $\mathcal{X}$ $(= \mathcal{Y}_1 + \mathcal{Y}_2)$: que vaut $E\{\mathcal{X}|\mathcal{Y}_1 \in \{2,3,4,5,6\}\}$?

  1. On a $E\{\mathcal{X}|\mathcal{Y}_1=1\} = E\{\mathcal{Y}_1|\mathcal{Y}_1=1\} + E\{\mathcal{Y}_2|\mathcal{Y}_1=1\} = 1 + E\{\mathcal{Y}_2\} = 4.5$.
  2. On a $E\{\mathcal{X}\} = P_{\mathcal{Y}_1}(1)E\{\mathcal{X}|\mathcal{Y}_1=1\} + P_{\mathcal{Y}_1}(\{2,3,4,5,6\})E\{\mathcal{X}|\mathcal{Y}_1 \in \{2,3,4,5,6\}\}$.

  Donc, $E\{\mathcal{X}|\mathcal{Y}_1 \in \{2,3,4,5,6\}\} = \frac{E\{\mathcal{X}\} - P_{\mathcal{Y}_1}(1)E\{\mathcal{X}|\mathcal{Y}_1=1\}}{P_{\mathcal{Y}_1}(\{2,3,4,5,6\})} = \frac{7-(1/6)4.5}{5/6} = 7.5$

- On peut aussi se convaincre que

$$E\{\mathcal{X}\} = \sum_{i=1}^{6} P_{\mathcal{Y}_1}(i)E\{\mathcal{X}|\mathcal{Y}_1=i\} = \frac{4.5+5.5+6.5+7.5+8.5+9.5}{6}.$$

- Nous reviendrons sur l'étude de l'espérance conditionnelle au chapitre 4.

Inégalité de Markov

Inégalité de Markov

Si $\mathcal{X}$ est une variable aléatoire positive (ou nulle) et d'espérance $\mu_{\mathcal{X}}$ finie (et donc aussi positive), alors $\forall c > 0$ on a

$$P(\mathcal{X} \geq c\mu_{\mathcal{X}}) \leq \frac{1}{c}. \tag{17}$$

Cette inégalité nous indique qu'une variable aléatoire positive ne peut dévier très au dessus de son espérance que très rarement.

Exemple. Si la durée de vie moyenne d'une batterie de voiture est de 3 ans, alors au moins 50% des batteries de voiture auront cessé de fonctionner après 6 ans, et au moins 75% auront cessé de fonctionner après 12 ans.

Exemple (bis). Soit $A \in \mathcal{E}$, et considérons la v.a. 1_A. On a $\mu_{1_A} = P(A)$. Prenons $c = 1/P(A)$. L'inégalité de Markov nous dit que $P(1_A \geq 1) \leq P(A)$. En fait ici on a $P(1_A \geq 1) = P(A)$. Cet exemple montre donc que la borne supérieure fournie par l'inégalité de Markov n'est pas améliorable en général.

Remarque. Dans la suite nous discuterons une autre inégalité, celle de Bienaymé-Tchebyshev, qui fait intervenir la notion d'écart-type, et qui est une conséquence directe de l'inégalité de Markov.

Démonstration de l'inégalité de Markov

- Remarquons tout d'abord que, l'inégalité de Markov est équivalente à dire que si $\mathcal{X} \geq 0$ alors $\forall c > 0$ on a

$$cP(\mathcal{X}(\omega) \geq c) \leq E\{\mathcal{X}\}.$$

Démonstration de l'inégalité de Markov

- Remarquons tout d'abord que, l'inégalité de Markov est équivalente à dire que si $\mathcal{X} \geq 0$ alors $\forall c > 0$ on a
$$cP(\mathcal{X}(\omega) \geq c) \leq E\{\mathcal{X}\}.$$
- Considérons alors l'événement $A = \{\omega : \mathcal{X}(\omega) \geq c\}$, et considérons la v.a. non-négative simple $\mathcal{Z}(\omega) = c1_A(\omega)$.

Démonstration de l'inégalité de Markov

- Remarquons tout d'abord que, l'inégalité de Markov est équivalente à dire que si $\mathcal{X} \geq 0$ alors $\forall c > 0$ on a
$$cP(\mathcal{X}(\omega) \geq c) \leq E\{\mathcal{X}\}.$$
- Considérons alors l'événement $A = \{\omega : \mathcal{X}(\omega) \geq c\}$, et considérons la v.a. non-négative simple $\mathcal{Z}(\omega) = c1_A(\omega)$.
- Par définition de l'espérance d'une v.a. non-négative simple on a
$$E\{\mathcal{Z}\} = cP(A) = cP(\mathcal{X}(\omega) \geq c).$$

Démonstration de l'inégalité de Markov

- Remarquons tout d'abord que, l'inégalité de Markov est équivalente à dire que si $\mathcal{X} \geq 0$ alors $\forall c > 0$ on a
$$cP(\mathcal{X}(\omega) \geq c) \leq E\{\mathcal{X}\}.$$
- Considérons alors l'événement $A = \{\omega : \mathcal{X}(\omega) \geq c\}$, et considérons la v.a. non-négative simple $\mathcal{Z}(\omega) = c1_A(\omega)$.
- Par définition de l'espérance d'une v.a. non-négative simple on a
$$E\{\mathcal{Z}\} = cP(A) = cP(\mathcal{X}(\omega) \geq c).$$
- Montrons que $\forall \omega \in \Omega : \mathcal{Z}(\omega) \leq \mathcal{X}(\omega)$. En effet
  - soit $\omega \in A$, et on a $\mathcal{Z}(\omega) = c$ et $\mathcal{X}(\omega) \geq c$ (par définition de A),
  - soit $\omega \notin A$, et on a $\mathcal{Z}(\omega) = 0$, et $\mathcal{X}(\omega) \geq 0$ (par hypothèse sur $\mathcal{X}$).

Démonstration de l'inégalité de Markov

- Remarquons tout d'abord que, l'inégalité de Markov est équivalente à dire que si $\mathcal{X} \geq 0$ alors $\forall c > 0$ on a
$$cP(\mathcal{X}(\omega) \geq c) \leq E\{\mathcal{X}\}.$$
- Considérons alors l'événement $A = \{\omega : \mathcal{X}(\omega) \geq c\}$, et considérons la v.a. non-négative simple $\mathcal{Z}(\omega) = c1_A(\omega)$.
- Par définition de l'espérance d'une v.a. non-négative simple on a
$$E\{\mathcal{Z}\} = cP(A) = cP(\mathcal{X}(\omega) \geq c).$$
- Montrons que $\forall \omega \in \Omega : \mathcal{Z}(\omega) \leq \mathcal{X}(\omega)$. En effet
  - soit $\omega \in A$, et on a $\mathcal{Z}(\omega) = c$ et $\mathcal{X}(\omega) \geq c$ (par définition de A),
  - soit $\omega \notin A$, et on a $\mathcal{Z}(\omega) = 0$, et $\mathcal{X}(\omega) \geq 0$ (par hypothèse sur $\mathcal{X}$).
- Comme $\mathcal{Z} \leq \mathcal{X}$ on a $E\{\mathcal{Z}\} \leq E\{\mathcal{X}\}$, étant donnée la définition de l'espérance d'une v.a. positive quelconque (cf. équation (14)).

Démonstration de l'inégalité de Markov

- Remarquons tout d'abord que, l'inégalité de Markov est équivalente à dire que si $\mathcal{X} \geq 0$ alors $\forall c > 0$ on a
$$cP(\mathcal{X}(\omega) \geq c) \leq E\{\mathcal{X}\}.$$
- Considérons alors l'événement $A = \{\omega : \mathcal{X}(\omega) \geq c\}$, et considérons la v.a. non-négative simple $\mathcal{Z}(\omega) = c1_A(\omega)$.
- Par définition de l'espérance d'une v.a. non-négative simple on a
$$E\{\mathcal{Z}\} = cP(A) = cP(\mathcal{X}(\omega) \geq c).$$
- Montrons que $\forall \omega \in \Omega : \mathcal{Z}(\omega) \leq \mathcal{X}(\omega)$. En effet
  - soit $\omega \in A$, et on a $\mathcal{Z}(\omega) = c$ et $\mathcal{X}(\omega) \geq c$ (par définition de A),
  - soit $\omega \notin A$, et on a $\mathcal{Z}(\omega) = 0$, et $\mathcal{X}(\omega) \geq 0$ (par hypothèse sur $\mathcal{X}$).
- Comme $\mathcal{Z} \leq \mathcal{X}$ on a $E\{\mathcal{Z}\} \leq E\{\mathcal{X}\}$, étant donnée la définition de l'espérance d'une v.a. positive quelconque (cf. équation (14)).
- c.q.f.d.

Espérance d'une fonction d'une v.a.

Espérance d'une fonction à valeurs réelles d'une v.a. discrète

$$E\{\phi(\mathcal{X})\} = \sum_{k=1}^{\infty} \phi(x_k) P(\mathcal{X} = x_k) = \sum_{k=1}^{\infty} \phi(x_k) P_{\mathcal{X}}(x_k), \tag{18}$$

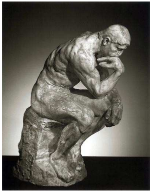
Espérance d'une fonction à valeurs réelles d'une v.a. réelle continue

$$E\{\phi(\mathcal{X})\} = \int_{\mathbb{R}} \phi(x) f_{\mathcal{X}}(x)\, \mathrm{d}x, \tag{19}$$

Ces formules découlent de la définition de l'espérance, pour autant que les séries/intégrales convergent absolument.

1. Fonction constante ($\phi(x) = a$) : $E\{\phi\} = a$.
2. Produit par une constante ($\phi(x) = ax$) : $E\{\phi\} = aE\{\mathcal{X}\}$.
3. Somme avec une constante ($\phi(x) = x + a$) : $E\{\phi\} = E\{\mathcal{X}\} + a$.
4. Fonction additive ($\phi(x) = \sum_{i=1}^{n} \phi_i(x)$) : $E\{\phi\} = \sum_{i=1}^{n} E\{\phi_i\}$.

Fonctions convexes et inégalité de Jensen



Homework : lire la section 3.5.4.1

Espérance d'une fonction de plusieurs v.a. (a)

Espérance mathématique d'une fonction $\phi(\mathcal{X},\mathcal{Y})$ de deux v.a. discrètes

Si les deux variables $\mathcal{X}$ et $\mathcal{Y}$ sont discrètes, le couple $\mathcal{Z} = (\mathcal{X},\mathcal{Y})$ est encore une variable aléatoire discrète et l'espérance de la variable ϕ se déduit donc de ce qui précède, et est donnée par la formule

$$E\{\phi(\mathcal{X},\mathcal{Y})\} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\phi(x_i,y_j)P_{\mathcal{X},\mathcal{Y}}(x_i,y_j), \tag{20}$$

où $P_{\mathcal{X},\mathcal{Y}}(x_i,y_j)$ désigne $P(\mathcal{X} = x_i \wedge \mathcal{Y} = y_j)$, et pour autant que la série converge absolument.

Notons que si $\Omega_{\mathcal{X}}$ et $\Omega_{\mathcal{Y}}$ sont dénombrables, il en est de même de leur produit cartésien $\Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$, et donc le couple $\mathcal{Z} = (\mathcal{X},\mathcal{Y})$ est une variable aléatoire discrète dans ces conditions.

Espérance d'une fonction de plusieurs v.a. (b)

Espérance mathématique d'une fonction $\phi(\mathcal{X}, \mathcal{Y})$ de deux v.a. conjointement continues

Si les deux variables sont continues et possèdent une densité conjointe (voir chapitre 4), l'espérance de la variable ϕ est donnée par la formule

$$E\{\phi(\mathcal{X}, \mathcal{Y})\} = \int_{\mathbb{R}} \int_{\mathbb{R}} \phi(x, y) f_{\mathcal{X},\mathcal{Y}}(x, y)\, \mathrm{d}x\, \mathrm{d}y, \tag{21}$$

pour autant que l'intégrale double converge.

Notons que si la fonction ϕ ne dépend en réalité que de l'une des deux variables aléatoires (disons qu'elle peut s'écrire sous la forme $\phi(\mathcal{X})$), alors ces deux formules donnent respectivement (voir chapitre 4)

$$E\{\phi(\mathcal{X}, \mathcal{Y})\} = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \phi(x_i) P_{\mathcal{X},\mathcal{Y}}(x_i, x_j) = \sum_{i=1}^{\infty} \phi(x_i) \left(\sum_{j=1}^{\infty} P_{\mathcal{X},\mathcal{Y}}(x_i, x_j) \right) = \sum_{i=1}^{\infty} \phi(x_i) P_{\mathcal{X}}(x_i)$$

et

$$E\{\phi(\mathcal{X}, \mathcal{Y})\} = \int_{\mathbb{R}} \int_{\mathbb{R}} \phi(x) f_{\mathcal{X},\mathcal{Y}}(x, y)\, \mathrm{d}x\, \mathrm{d}y = \int_{\mathbb{R}} \phi(x) \left(\int_{\mathbb{R}} f_{\mathcal{X},\mathcal{Y}}(x, y)\, \mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}} \phi(x) f_{\mathcal{X}}(x)\, \mathrm{d}x,$$

c'est-à-dire les formules déjà présentées plus haut.

Espérance d'une fonction de plusieurs v.a. (c)

Par ailleurs, si la fonction ϕ est une somme de deux (ou plusieurs) fonctions, son espérance est la somme des espérances de ces fonctions (pour autant que les espérances en question soient finies), toujours à cause de la linéarité de l'opérateur d'intégration/sommation:

$$E\{\sum_{i=1}^{n} \phi_i(\mathcal{X},\mathcal{Y})\} = \sum_{i=1}^{n} E\{\phi_i(\mathcal{X},\mathcal{Y})\}. \tag{22}$$

Enfin, ces idées peuvent être généralisées, en considérant un nombre fini quelconque de variables aléatoires $\mathcal{X}_i$ et un nombre quelconque fini de fonctions ϕ_j de ces variables aléatoires, de la manière suivante:

$$E\{\sum_{i=1}^{n} \phi_i(\mathcal{X}_1,\ldots,\mathcal{X}_m)\} = \sum_{i=1}^{n} E\{\phi_i(\mathcal{X}_1,\ldots,\mathcal{X}_m)\}. \tag{23}$$

Espérance d'une fonction de plusieurs v.a. (d)

On en déduit la propriété extrêmement importante suivant :

Linéarité de l'espérance mathématique

$$E\{\sum_{i=1}^{n} \lambda_i \mathcal{X}_i)\} = \sum_{i=1}^{n} \lambda_i E\{\mathcal{X}_i\}, \tag{24}$$

qui exprime le fait que l'espérance mathématique d'une combinaison linéaire de variables aléatoires d'espérance finie est la combinaison linéaire correspondante des espérances mathématiques de ces variables
(**sans hypothèse d'indépendance entre les variables aléatoires en question**).

Ce théorème reste vrai même si certaines des variables aléatoires sont ni continues ni discrètes, pour autant qu'elles soient toutes d'espérance finie.

Espérance d'une fonction de plusieurs v.a. (e)

Espérance mathématique d'un produit de deux variables aléatoires

Définissant une fonction ϕ, produit de deux v.a. $(\phi(\mathcal{X},\mathcal{Y}) = \mathcal{X}\mathcal{Y})$, son espérance est définie de manière générique par

$$E\{\mathcal{X}\mathcal{Y}\} = \int_{\mathbb{R}^2} xy \,\mathrm{d}P_{\mathcal{X}\mathcal{Y}}(x,y).$$

(Suggestion : écrire cette formule lorsque les variables $\mathcal{X},\mathcal{Y}$ sont conjointement continues, ou bien discrètes.)

Lorsque $\mathcal{X}$ et $\mathcal{Y}$ sont indépendantes, la mesure $P_{\mathcal{X}\mathcal{Y}}(x,y)$ se factorise et donc l'intégrale double peut se décomposer en un produit de deux intégrales simples :

$$E\{\mathcal{X}\mathcal{Y}\} = \int_{\mathbb{R}} x \,\mathrm{d}P_{\mathcal{X}}(x) \int_{\mathbb{R}} y \,\mathrm{d}P_{\mathcal{Y}}(y) = E\{\mathcal{X}\}E\{\mathcal{Y}\}.$$

La réciproque n'est pas vraie.

NB. Les travaux pratiques et séances de répétition illustreront ces idées.

3.6 Variance, écart-type, covariance : définitions

Lorsque l'espérance $\mu_{\mathcal{X}} = E\{\mathcal{X}\}$ est finie, la **variance** de la variable aléatoire $\mathcal{X}$ est définie par

$$V\{\mathcal{X}\} = \sigma^2_{\mathcal{X}} = E\{(\mathcal{X} - \mu_{\mathcal{X}})^2\}, \tag{25}$$

lorsque cette grandeur est finie.

L'**écart-type**, désigné par $\sigma_{\mathcal{X}}$, est la racine carrée positive de la variance $V\{\mathcal{X}\}$.

On définit la **covariance** de deux variables aléatoires réelles, $\mathcal{X}$ et $\mathcal{Y}$, par

$$\mathrm{cov}\{\mathcal{X};\mathcal{Y}\} \triangleq E\left\{(\mathcal{X} - E\{\mathcal{X}\})(\mathcal{Y} - E\{\mathcal{Y}\})\right\} = E\{\mathcal{X}\mathcal{Y}\} - E\{\mathcal{X}\}E\{\mathcal{Y}\}, \tag{26}$$

lorsque cette grandeur est finie.

Notons que si $E\{\mathcal{X}^2\}$ est finie, alors l'espérance $E\{\mathcal{X}\}$ l'est aussi et aussi la variance $V\{\mathcal{X}\}$.
Si $E\{\mathcal{X}^2\}$ et $E\{\mathcal{Y}^2\}$ sont finies alors $\mathrm{cov}\{\mathcal{X};\mathcal{Y}\}$ l'est aussi.
Nous reviendrons au chapitre 4 sur les conditions d'existence et les relations entre ces grandeurs.

3.6 Variance, écart-type, covariance : illustration

- $\mathcal{Y}_1$: $\mu_{\mathcal{Y}_1} = 3.5$, et $V\{\mathcal{Y}_1\} = E\{(\mathcal{Y}_1 - \mu_{\mathcal{Y}_1})^2\} = \sum_{y_1=1}^{6}(y_1 - 3.5)^2 P_{\mathcal{Y}_1}(y_1)$
  Pour rappel: $P_{\mathcal{Y}_1}(y_1) = 1/6, \forall y_1$
  On trouve : $V\{\mathcal{Y}_1\} = 35/12 = 2.91666667$, et $\sigma_{\mathcal{Y}_1} = 1.707825128$.

  NB: on aurait aussi pu calculer $V\{\mathcal{Y}_1\} = \sum_{\omega\in\Omega}(\mathcal{Y}_1(\omega) - 3.5)^2 P_\Omega(\omega)$, avec $P_\Omega(\omega) = 1/36, \forall\omega$.

3.6 Variance, écart-type, covariance : illustration

- $\mathcal{Y}_1$: $\mu_{\mathcal{Y}_1} = 3.5$, et $V\{\mathcal{Y}_1\} = E\{(\mathcal{Y}_1 - \mu_{\mathcal{Y}_1})^2\} = \sum_{y_1=1}^{6}(y_1 - 3.5)^2 P_{\mathcal{Y}_1}(y_1)$
  Pour rappel: $P_{\mathcal{Y}_1}(y_1) = 1/6, \forall y_1$
  On trouve : $V\{\mathcal{Y}_1\} = 35/12 = 2.91666667$, et $\sigma_{\mathcal{Y}_1} = 1.707825128$.
  NB: on aurait aussi pu calculer $V\{\mathcal{Y}_1\} = \sum_{\omega\in\Omega}(\mathcal{Y}_1(\omega) - 3.5)^2 P_\Omega(\omega)$, avec $P_\Omega(\omega) = 1/36, \forall\omega$.
- $\mathcal{Y}_2$: $V\{\mathcal{Y}_2\} = 2.91666667$, et $\sigma_{\mathcal{Y}_2} = 1.707825128$, puisque cette variable est distribuée identiquement à $\mathcal{Y}_1$.

3.6 Variance, écart-type, covariance : illustration

- $\mathcal{Y}_1$: $\mu_{\mathcal{Y}_1} = 3.5$, et $V\{\mathcal{Y}_1\} = E\{(\mathcal{Y}_1 - \mu_{\mathcal{Y}_1})^2\} = \sum_{y_1=1}^{6}(y_1 - 3.5)^2 P_{\mathcal{Y}_1}(y_1)$
  Pour rappel: $P_{\mathcal{Y}_1}(y_1) = 1/6, \forall y_1$
  On trouve : $V\{\mathcal{Y}_1\} = 35/12 = 2.91666667$, et $\sigma_{\mathcal{Y}_1} = 1.707825128$.

  NB: on aurait aussi pu calculer $V\{\mathcal{Y}_1\} = \sum_{\omega\in\Omega}(\mathcal{Y}_1(\omega) - 3.5)^2 P_\Omega(\omega)$, avec $P_\Omega(\omega) = 1/36, \forall\omega$.
- $\mathcal{Y}_2$: $V\{\mathcal{Y}_2\} = 2.91666667$, et $\sigma_{\mathcal{Y}_2} = 1.707825128$, puisque cette variable est distribuée identiquement à $\mathcal{Y}_1$.
- En ce qui concerne la variable $\mathcal{X} = \mathcal{Y}_1 + \mathcal{Y}_2$ on a $\mu_{\mathcal{X}} = \mu_{\mathcal{Y}_1} + \mu_{\mathcal{Y}_2} = 7$, et donc $V\{\mathcal{X}\} = \sum_{x=2}^{12}(x-7)^2 P_{\mathcal{X}}(x)$.
  Voir ce qui précède pour les valeurs de $P_{\mathcal{X}}(x)$
  On trouve : $V\{\mathcal{X}\} = 35/6 = 5.83333333$, et $\sigma_{\mathcal{X}} = 2.415229458$.
  NB: on aurait aussi pu calculer $V\{\mathcal{X}\} = \sum_{y_1=1}^{6}\sum_{y_2=1}^{6}(y_1+y_2-7)(y_1+y_2-7)P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1,y_2)$!

3.6 Variance, écart-type, covariance : illustration

- $\mathcal{Y}_1$: $\mu_{\mathcal{Y}_1} = 3.5$, et $V\{\mathcal{Y}_1\} = E\{(\mathcal{Y}_1 - \mu_{\mathcal{Y}_1})^2\} = \sum_{y_1=1}^{6}(y_1 - 3.5)^2 P_{\mathcal{Y}_1}(y_1)$
  Pour rappel: $P_{\mathcal{Y}_1}(y_1) = 1/6, \forall y_1$
  On trouve : $V\{\mathcal{Y}_1\} = 35/12 = 2.91666667$, et $\sigma_{\mathcal{Y}_1} = 1.707825128$.
  NB: on aurait aussi pu calculer $V\{\mathcal{Y}_1\} = \sum_{\omega \in \Omega}(\mathcal{Y}_1(\omega) - 3.5)^2 P_\Omega(\omega)$, avec $P_\Omega(\omega) = 1/36, \forall \omega$.
- $\mathcal{Y}_2$: $V\{\mathcal{Y}_2\} = 2.91666667$, et $\sigma_{\mathcal{Y}_2} = 1.707825128$, puisque cette variable est distribuée identiquement à $\mathcal{Y}_1$.
- En ce qui concerne la variable $\mathcal{X} = \mathcal{Y}_1 + \mathcal{Y}_2$ on a $\mu_{\mathcal{X}} = \mu_{\mathcal{Y}_1} + \mu_{\mathcal{Y}_2} = 7$, et donc $V\{\mathcal{X}\} = \sum_{x=2}^{12}(x - 7)^2 P_{\mathcal{X}}(x)$.
  Voir ce qui précède pour les valeurs de $P_{\mathcal{X}}(x)$
  On trouve : $V\{\mathcal{X}\} = 35/6 = 5.83333333$, et $\sigma_{\mathcal{X}} = 2.415229458$.
  NB: on aurait aussi pu calculer $V\{\mathcal{X}\} = \sum_{y_1=1}^{6}\sum_{y_2=1}^{6}(y_1 + y_2 - 7)(y_1 + y_2 - 7)P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1, y_2)$!
- $\mathrm{cov}\{\mathcal{Y}_1; \mathcal{Y}_2\} = \sum_{y_1=1}^{6}\sum_{y_2=1}^{6}(y_1 - 3.5)(y_2 - 3.5)P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1, y_2)$
  Sachant que $P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1, y_2) = 1/36, \forall y_1, y_2$ on trouve $\mathrm{cov}\{\mathcal{Y}_1; \mathcal{Y}_2\} = 0$.

3.6 Variance, écart-type, covariance : illustration

- $\mathcal{Y}_1$: $\mu_{\mathcal{Y}_1} = 3.5$, et $V\{\mathcal{Y}_1\} = E\{(\mathcal{Y}_1 - \mu_{\mathcal{Y}_1})^2\} = \sum_{y_1=1}^{6}(y_1 - 3.5)^2 P_{\mathcal{Y}_1}(y_1)$
  Pour rappel: $P_{\mathcal{Y}_1}(y_1) = 1/6, \forall y_1$
  On trouve : $V\{\mathcal{Y}_1\} = 35/12 = 2.91666667$, et $\sigma_{\mathcal{Y}_1} = 1.707825128$.
  NB: on aurait aussi pu calculer $V\{\mathcal{Y}_1\} = \sum_{\omega\in\Omega}(\mathcal{Y}_1(\omega) - 3.5)^2 P_\Omega(\omega)$, avec $P_\Omega(\omega) = 1/36, \forall\omega$.
- $\mathcal{Y}_2$: $V\{\mathcal{Y}_2\} = 2.91666667$, et $\sigma_{\mathcal{Y}_2} = 1.707825128$, puisque cette variable est distribuée identiquement à $\mathcal{Y}_1$.
- En ce qui concerne la variable $\mathcal{X} = \mathcal{Y}_1 + \mathcal{Y}_2$ on a $\mu_{\mathcal{X}} = \mu_{\mathcal{Y}_1} + \mu_{\mathcal{Y}_2} = 7$, et donc $V\{\mathcal{X}\} = \sum_{x=2}^{12}(x-7)^2 P_{\mathcal{X}}(x)$.
  Voir ce qui précède pour les valeurs de $P_{\mathcal{X}}(x)$
  On trouve : $V\{\mathcal{X}\} = 35/6 = 5.83333333$, et $\sigma_{\mathcal{X}} = 2.415229458$.
  NB: on aurait aussi pu calculer $V\{\mathcal{X}\} = \sum_{y_1=1}^{6}\sum_{y_2=1}^{6}(y_1+y_2-7)(y_1+y_2-7)P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1,y_2)$!
- $\mathrm{cov}\{\mathcal{Y}_1;\mathcal{Y}_2\} = \sum_{y_1=1}^{6}\sum_{y_2=1}^{6}(y_1-3.5)(y_2-3.5)P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1,y_2)$
  Sachant que $P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1,y_2) = 1/36, \forall y_1, y_2$ on trouve $\mathrm{cov}\{\mathcal{Y}_1;\mathcal{Y}_2\} = 0$.
- $\mathrm{cov}\{\mathcal{Y}_1;\mathcal{X}\} = \sum_{y_1=1}^{6}\sum_{y_2=1}^{6}(y_1-3.5)(y_1+y_2-7)P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1,y_2)$
  On trouve que $\mathrm{cov}\{\mathcal{Y}_1;\mathcal{X}\} = 2.91666667$.

3.6 Variance et covariance : propriétés de base (a)

On a (a étant un nombre réel constant, et $\mu_{\mathcal{X}}$ désigant $E\{\mathcal{X}\}$)

$$E\{(\mathcal{X}-a)^2\} = V\{\mathcal{X}\} + (\mu_{\mathcal{X}} - a)^2. \tag{27}$$

En effet,

$$\begin{aligned}
E\{(\mathcal{X}-a)^2\} &= E\{(\mathcal{X}-\mu_{\mathcal{X}}+\mu_{\mathcal{X}}-a)^2\} && (28)\\
&= E\{(\mathcal{X}-\mu_{\mathcal{X}})^2 + 2(\mathcal{X}-\mu_{\mathcal{X}})(\mu_{\mathcal{X}}-a) + (\mu_{\mathcal{X}}-a)^2\} && (29)\\
&= E\{(\mathcal{X}-\mu_{\mathcal{X}})^2\} + E\{2(\mathcal{X}-\mu_{\mathcal{X}})(\mu_{\mathcal{X}}-a)\} + E\{(\mu_{\mathcal{X}}-a)^2\} && (30)\\
&= V\{\mathcal{X}\} + 0 + (\mu_{\mathcal{X}}-a)^2, && (31)
\end{aligned}$$

puisque $(\mu_{\mathcal{X}} - a)^2$ est une constante, et que $E\{(\mathcal{X} - \mu_{\mathcal{X}})\} = 0$.

Par conséquent, $V\{\mathcal{X}\} \leq E\{(\mathcal{X}-a)^2\}, \forall a \in \mathbb{R}$, et la valeur de $a = E\{\mathcal{X}\} = \mu_{\mathcal{X}}$ minimise donc l'écart quadratique moyen $E\{(\mathcal{X}-a)^2\}$.

On dit que que $E\{\mathcal{X}\}$ est la constante qui est la plus proche au sens des moindres carrés de la variable aléatoire $\mathcal{X}$, et son écart quadratique moyen est mesuré par la variance de $\mathcal{X}$.

3.6 Variance et covariance : propriétés de base (b)

On en déduit de $E\{(\mathcal{X}-a)^2\} = V\{\mathcal{X}\} + (\mu_{\mathcal{X}} - a)^2$, en prenant $a = 0$ que

$$V\{\mathcal{X}\} = E\{X^2\} - (E\{\mathcal{X}\})^2. \tag{32}$$

Cette propriété est utile en pratique.

Par ailleurs, on a les propriétés suivantes :

- $V\{\mathcal{X} + a\} = V\{\mathcal{X}\}$.
  Puisque $(\mathcal{X} + a - E\{\mathcal{X} + a\})^2 = (\mathcal{X} + a - E\{\mathcal{X}\} - a)^2 = (\mathcal{X} - E\{\mathcal{X}\})^2$.
- $V\{a\mathcal{X}\} = a^2 V\{\mathcal{X}\}$.
  $V\{a\mathcal{X}\} = E\{(a\mathcal{X} - E\{a\mathcal{X}\})^2\} = E\{(a\mathcal{X} - aE\{\mathcal{X}\})^2\} = E\{a^2(\mathcal{X} - E\{\mathcal{X}\})^2\} = a^2 E\{(\mathcal{X} - E\{\mathcal{X}\})^2\}$.
- $V\{\mathcal{X} + \mathcal{Y}\} = V\{\mathcal{X}\} + V\{\mathcal{Y}\} + 2\text{cov}\{\mathcal{X};\mathcal{Y}\}$.
  Si les v.a. sont indépendantes, on a $E\{\mathcal{X}\mathcal{Y}\} = E\{\mathcal{X}\}E\{\mathcal{Y}\}$ et donc $\text{cov}\{\mathcal{X};\mathcal{Y}\} = 0$. Dans ce cas

$$V\{\mathcal{X} + \mathcal{Y}\} = V\{\mathcal{X}\} + V\{\mathcal{Y}\}.$$

La réciproque n'est pas vraie.

3.6 Variance, covariance : illustration

- $\mathcal{Y}_1 \perp \mathcal{Y}_2$: on en déduit que $\mathrm{cov}\{\mathcal{Y}_1; \mathcal{Y}_2\} = 0$ comme déjà calculé plus haut. En effet, on a

$$
\begin{aligned}
\mathrm{cov}\{\mathcal{Y}_1; \mathcal{Y}_2\} &= \sum_{y_1=1}^{6} \sum_{y_2=1}^{6} (y_1 - 3.5)(y_2 - 3.5) P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1, y_2) &\quad (33)\\
&= \sum_{y_1=1}^{6} \sum_{y_2=1}^{6} (y_1 - 3.5)(y_2 - 3.5) P_{\mathcal{Y}_1}(y_1) P_{\mathcal{Y}_2}(y_2) &\quad (34)\\
&= \left(\sum_{y_1=1}^{6} (y_1 - 3.5) P_{\mathcal{Y}_1}(y_1) \right) \left(\sum_{y_2=1}^{6} (y_2 - 3.5) P_{\mathcal{Y}_2}(y_2) \right) &\quad (35)\\
&= (E\{\mathcal{Y}_1\} - 3.5)\,(E\{\mathcal{Y}_2\} - 3.5) = 0. &\quad (36)
\end{aligned}
$$

3.6 Variance, covariance : illustration

- $\mathcal{Y}_1 \perp \mathcal{Y}_2$: on en déduit que $\mathrm{cov}\{\mathcal{Y}_1;\mathcal{Y}_2\} = 0$ comme déjà calculé plus haut. En effet, on a

$$\begin{aligned}
\mathrm{cov}\{\mathcal{Y}_1;\mathcal{Y}_2\} &= \sum_{y_1=1}^{6}\sum_{y_2=1}^{6}(y_1-3.5)(y_2-3.5)P_{\mathcal{Y}_1,\mathcal{Y}_2}(y_1,y_2) &(33)\\
&= \sum_{y_1=1}^{6}\sum_{y_2=1}^{6}(y_1-3.5)(y_2-3.5)P_{\mathcal{Y}_1}(y_1)P_{\mathcal{Y}_2}(y_2) &(34)\\
&= \left(\sum_{y_1=1}^{6}(y_1-3.5)P_{\mathcal{Y}_1}(y_1)\right)\left(\sum_{y_2=1}^{6}(y_2-3.5)P_{\mathcal{Y}_2}(y_2)\right) &(35)\\
&= (E\{\mathcal{Y}_1\}-3.5)\,(E\{\mathcal{Y}_2\}-3.5) = 0. &(36)
\end{aligned}$$

- On peut aussi vérifier cela autrement :

  $V\{\mathcal{X}\} = V\{\mathcal{Y}_1+\mathcal{Y}_2\} = V\{\mathcal{Y}_1\} + V\{\mathcal{Y}_2\} + \mathrm{cov}\{\mathcal{Y}_1;\mathcal{Y}_2\}$.
  Or, nous avions calculé que $V\{\mathcal{X}\} = 35/6$ et que $V\{\mathcal{Y}_1\} = V\{\mathcal{Y}_2\} = 35/12$.
  On doit donc avoir $\mathrm{cov}\{\mathcal{Y}_1;\mathcal{Y}_2\} = 0$.

3.6 Bienaymé-Tchebyshev

L'espérance et la variance sont reliées par *l'inégalité de Bienaymé-Tchebyshev* :

Inégalité de Bienaymé-Tchebyshev

Si $\mathcal{X}$ est une variable aléatoire d'espérance $\mu_{\mathcal{X}}$ et d'écart-type $\sigma_{\mathcal{X}}$, on a $\forall c > 0$:

$$P(|\mathcal{X} - \mu_{\mathcal{X}}| \geq c\sigma_{\mathcal{X}}) \leq \frac{1}{c^2}. \tag{37}$$

On déduit de cette inégalité que si $\sigma_{\mathcal{X}} = 0$ la v.a. est presque sûrement égale à son espérance $\mu_{\mathcal{X}}$, c'est-à-dire presque sûrement constante. La variance mesure donc bien le caractère aléatoire d'une v.a. du point de vue de ses écarts possibles par rapport à son espérance.

Notons que cette inégalité est une conséquence directe de l'inégalité de Markov appliquée à la variable aléatoire positive $\mathcal{Y} = (\mathcal{X} - \mu_{\mathcal{X}})^2$ dont l'espérance vaut $\sigma^2_{\mathcal{X}}$ et est donc finie lorsque la variance de $\mathcal{X}$ l'est.

Application de l'inégalité de Bienaymé-Tchebyshev

Loi des grand nombres pour n lancers successifs d'une pièce équilibrée.

- Soit $\mathcal{Z}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i$. On a (Rappel, les $\mathcal{X}_i$ sont i.i.d. et $P_{\mathcal{X}_i}(0) = P_{\mathcal{X}_i}(1) = 1/2, \forall i.$):
  - $\forall i, E\{\mathcal{X}_i\} = 0P_{\mathcal{X}_i}(0) + 1P_{\mathcal{X}_i}(1) = 1/2$, et
  - $\forall i, V\{\mathcal{X}_i\} = (0-1/2)^2 P_{\mathcal{X}_i}(0) + (1-1/2)^2 P_{\mathcal{X}_i}(1) = 1/4$.
  - et donc $\mu_{\mathcal{Z}_n} = E\{\mathcal{Z}_n\} = \frac{1}{n}\sum_{i=1}^{n} E\{\mathcal{X}_i\} = 1/2$,
  - ainsi que $\sigma^2_{\mathcal{Z}_n} = V\{\mathcal{Z}_n\} = \frac{1}{n^2}\sum_{i=1}^{n} V\{\mathcal{X}_i\} = \frac{1}{4n}$.

Application de l'inégalité de Bienaymé-Tchebyshev

Loi des grand nombres pour n lancers successifs d'une pièce équilibrée.

- Soit $\mathcal{Z}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i$. On a (Rappel, les $\mathcal{X}_i$ sont i.i.d. et $P_{\mathcal{X}_i}(0) = P_{\mathcal{X}_i}(1) = 1/2, \forall i.$):
  - $\forall i$, $E\{\mathcal{X}_i\} = 0P_{\mathcal{X}_i}(0) + 1P_{\mathcal{X}_i}(1) = 1/2$, et
  - $\forall i$, $V\{\mathcal{X}_i\} = (0-1/2)^2 P_{\mathcal{X}_i}(0) + (1-1/2)^2 P_{\mathcal{X}_i}(1) = 1/4$.
  - et donc $\mu_{\mathcal{Z}_n} = E\{\mathcal{Z}_n\} = \frac{1}{n}\sum_{i=1}^{n} E\{\mathcal{X}_i\} = 1/2$,
  - ainsi que $\sigma^2_{\mathcal{Z}_n} = V\{\mathcal{Z}_n\} = \frac{1}{n^2}\sum_{i=1}^{n} V\{\mathcal{X}_i\} = \frac{1}{4n}$.
- L'inégalité de B-T implique alors que $\forall \epsilon > 0$ on a
  - $P(|\mathcal{Z}_n - 1/2| \geq \epsilon) \leq \frac{\sigma^2_{\mathcal{Z}_n}}{\epsilon^2} = \frac{1}{4n\epsilon^2}$.

Application de l'inégalité de Bienaymé-Tchebyshev

Loi des grand nombres pour n lancers successifs d'une pièce équilibrée.

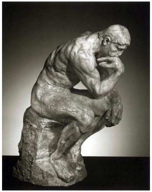
- Soit $\mathcal{Z}_n = \frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_i$. On a (Rappel, les $\mathcal{X}_i$ sont i.i.d. et $P_{\mathcal{X}_i}(0) = P_{\mathcal{X}_i}(1) = 1/2, \forall i.$):
  - $\forall i$, $E\{\mathcal{X}_i\} = 0P_{\mathcal{X}_i}(0) + 1P_{\mathcal{X}_i}(1) = 1/2$, et
  - $\forall i$, $V\{\mathcal{X}_i\} = (0-1/2)^2 P_{\mathcal{X}_i}(0) + (1-1/2)^2 P_{\mathcal{X}_i}(1) = 1/4$.
  - et donc $\mu_{\mathcal{Z}_n} = E\{\mathcal{Z}_n\} = \frac{1}{n}\sum_{i=1}^{n} E\{\mathcal{X}_i\} = 1/2$,
  - ainsi que $\sigma^2_{\mathcal{Z}_n} = V\{\mathcal{Z}_n\} = \frac{1}{n^2}\sum_{i=1}^{n} V\{\mathcal{X}_i\} = \frac{1}{4n}$.
- L'inégalité de B-T implique alors que $\forall\epsilon > 0$ on a
  - $P(|\mathcal{Z}_n - 1/2| \geq \epsilon) \leq \frac{\sigma^2_{\mathcal{Z}_n}}{\epsilon^2} = \frac{1}{4n\epsilon^2}$.
- Le nombre $N(n,\epsilon)$ de réalisations telles que $\mathcal{Z}_n \in [1/2-\epsilon, 1/2+\epsilon]$
  - vérifie $P(|\mathcal{Z}_n - 1/2| \leq \epsilon) = \frac{N(n,\epsilon)}{2^n}$.
  - On a donc $\frac{N(n,\epsilon)}{2^n} \geq 1 - P(|\mathcal{Z}_n - 1/2| \geq \epsilon) = 1 - \frac{1}{4n\epsilon^2}$.

Application de l'inégalité de Bienaymé-Tchebyshev

Loi des grand nombres pour n lancers successifs d'une pièce équilibrée.

- Soit $\mathcal{Z}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i$. On a (Rappel, les $\mathcal{X}_i$ sont i.i.d. et $P_{\mathcal{X}_i}(0) = P_{\mathcal{X}_i}(1) = 1/2, \forall i.$):
  - $\forall i, E\{\mathcal{X}_i\} = 0P_{\mathcal{X}_i}(0) + 1P_{\mathcal{X}_i}(1) = 1/2$, et
  - $\forall i, V\{\mathcal{X}_i\} = (0 - 1/2)^2 P_{\mathcal{X}_i}(0) + (1 - 1/2)^2 P_{\mathcal{X}_i}(1) = 1/4$.
  - et donc $\mu_{\mathcal{Z}_n} = E\{\mathcal{Z}_n\} = \frac{1}{n}\sum_{i=1}^{n} E\{\mathcal{X}_i\} = 1/2$,
  - ainsi que $\sigma^2_{\mathcal{Z}_n} = V\{\mathcal{Z}_n\} = \frac{1}{n^2}\sum_{i=1}^{n} V\{\mathcal{X}_i\} = \frac{1}{4n}$.
- L'inégalité de B-T implique alors que $\forall\epsilon > 0$ on a
  - $P(|\mathcal{Z}_n - 1/2| \geq \epsilon) \leq \frac{\sigma^2_{\mathcal{Z}_n}}{\epsilon^2} = \frac{1}{4n\epsilon^2}$.
- Le nombre $N(n, \epsilon)$ de réalisations telles que $\mathcal{Z}_n \in [1/2 - \epsilon, 1/2 + \epsilon]$
  - vérifie $P(|\mathcal{Z}_n - 1/2| \leq \epsilon) = \frac{N(n,\epsilon)}{2^n}$.
  - On a donc $\frac{N(n,\epsilon)}{2^n} \geq 1 - P(|\mathcal{Z}_n - 1/2| \geq \epsilon) = 1 - \frac{1}{4n\epsilon^2}$.
- On a donc $\forall\epsilon > 0 : \lim_{n\to\infty} \frac{N(n,\epsilon)}{2^n} = 1$.

Démonstration de l'inégalité de Bienaymé-Tchebyshev



Homework : faire la démonstration soi-même

Suggestion: appliquer l'inégalité de Markov à la v.a. $\mathcal{Y} = (\mathcal{X} - \mu_{\mathcal{X}})^2$, en vérifiant toutefois qu'elle peut bien s'y appliquer.

3.7 Autres moments

On définit, s'ils existent, les moments *centrés* d'ordre k par

$$\mu_k = E\left\{(\mathcal{X} - \mu_{\mathcal{X}})^k\right\}. \tag{38}$$

On a évidemment $\mu_1 = 0$ et $\mu_2 = \sigma^2_{\mathcal{X}}$. Si la distribution de la variable est symétrique par rapport à sa moyenne on a $\mu_{2k+1} = 0, \forall k$.

Le moment *non-centré* d'ordre k est quant à lui simplement défini par $E\left\{\mathcal{X}^k\right\}$. La moyenne est donc le moment non-centré d'ordre 1.

Les moments caractérisent ensemble (sauf exception rare) la loi de probabilité de la variable aléatoire (cf la discussion sur les fonctions caractéristiques et génératrices dans les sections qui suivent).

3.8 Lois de probabilité d'usage courant

Nous passons rapidement en revue les principale lois d'usage courant en insistant sur leurs propriétés mathématiques.

- Lois de variables discrètes
  - Uniforme
  - Bernoulli
  - Binomiale
  - Poisson
- Lois de variables continues
  - Uniforme
  - Exponentielle
  - Gaussienne
  - Cauchy

Remarque: lire attentivement la section 3.8 du syllabus pour des explications plus détaillées et une discussion des contextes applicatifs où ces modèles sont utilisés.

Loi uniforme discrète

Il s'agit de la loi d'une variable aléatoire $\mathcal{X}$ définie sur $\{1, 2, \ldots, n\}$ et telle que chacune de ses n valeurs possibles $i \in \{1, 2, \ldots, n\}$ soit de probabilité $P(\mathcal{X}(\omega) = i) = \frac{1}{n}$. On a donc

$$E\{\mathcal{X}\} = \sum_{i=1}^{n} i \frac{1}{n} = \frac{n+1}{2}$$

et

$$V\{\mathcal{X}\} = \sum_{i=1}^{n} \left(i - \frac{n+1}{2}\right)^2 \frac{1}{n} = \frac{n^2-1}{12}.$$

On peut en déduire que l'espérance d'une v.a. uniforme sur $\{m+1, 2, \ldots, m+n\}$ vaut $m + (n+1)/2$ et que sa variance reste $(n^2-1)/12$.

Suggestion: appliquer ces formules aux variables $\mathcal{Y}_i$ du double lancer de dé et aux variables $\mathcal{X}_i$ du lancer de n pièces.

Loi de Bernoulli

C'est la loi d'une v.a. $\mathcal{X}$ ne pouvant prendre que deux valeurs possibles 1 ou 0, avec les probabilités p et $1-p$. En d'autres termes, $\mathcal{X}$ est la fonction indicatrice d'un événement A de probabilité $P(A)=p$. On a

$$E\{\mathcal{X}\} = 1 * p + 0 * (1-p) = p$$

et

$$V\{\mathcal{X}\} = (1-p)^2 p + (0-p)^2 (1-p) = p(1-p).$$

Exemple 3 (suite). Consommation électrique d'un bâtiment

Supposons qu'un appareil (disons une lampe) ait une probabilité $p = 10^{-1}$ d'être branché à un moment particulier, et consomme à ce moment une puissance fixe (disons de $100W$). La consommation de cet appareil peut alors être modélisée par une variable $\mathcal{X}$ qui est le produit de la constante 100 et d'une variable de Bernoulli. On a donc que $E\{\mathcal{X}\} = 100p = 10W$, et que $V\{\mathcal{X}\} = 100^2 p(1-p) = 900$. Donc $\sigma_{\mathcal{X}} = 30W$.

Loi binomiale (voir répétitions, pour illustrations)

On répète n fois une expérience de Bernoulli, et on désigne par $\mathcal{X}$ la variable aléatoire qui compte le nombre de fois sur n que l'événement A est réalisé. $\mathcal{X}$ est la somme de n v.a. indépendantes et identiquement distribuées (i.i.d.) $\mathcal{X} = \sum_{i=1}^{n} \mathcal{X}_i$.

La loi de $\mathcal{X}$ est par définition la loi binomiale $\mathcal{B}(n, p)$. Les valeurs possibles de $\mathcal{X}$ sont $\{0, 1, \ldots, n\}$ On a $E\{\mathcal{X}\} = np$, et $V\{\mathcal{X}\} = np(1-p)$. D'autre part, on a

$$P(\mathcal{X} = k) = C_n^k p^k (1-p)^{(n-k)}.$$

On a la propriété importante (et évidente) suivante : soient $\mathcal{X} \sim \mathcal{B}(n_1, p)$ et $\mathcal{Y} \sim \mathcal{B}(n_2, p)$ indépendantes, alors $\mathcal{Z} = \mathcal{X} + \mathcal{Y} \sim \mathcal{B}(n_1 + n_2, p)$.

Remarque : si au lieu de compter le nombre total de succès, on considère la proportion de succès $\mathcal{Z} = \mathcal{X}/n$, on déduit que $E\{\mathcal{Z}\} = p$ et $V\{\mathcal{Z}\} = p(1-p)/n$. On constate que la variance de cette variable tend vers zéro lorsque le nombre d'essais tend vers l'infini. L'inégalité de Bienaymé-Tchebyshev permet donc d'affirmer que la proportion $\mathcal{Z}$ de succès tend vers la probabilité de succès p lorsque n est suffisamment grand.

Loi de Poisson

La loi de Poisson $\mathcal{P}(\lambda)$ est la loi d'une v.a. entière positive ou nulle décrite par

$$P_{\mathcal{X}}(x) = \exp(-\lambda)\frac{\lambda^x}{x!}.$$

On a $E\{\mathcal{X}\} = \lambda$, et $V\{\mathcal{X}\} = \lambda$.

On montre que si $\mathcal{X}_n \sim \mathcal{B}(n, p)$ est une suite de v.a. binomiales telle que

$$\lim_{n\to\infty} np = \lambda,$$

alors $\mathcal{X}_n$ converge en loi vers $\mathcal{P}(\lambda)$.

Exemple 3 (suite). Consommation électrique d'un bâtiment

Dans notre exemple ci-dessus on a $np = 10$, $p = 10^{-1}$, $n = 100$. On demande de comparer la loi $\mathcal{B}(n, p)$ avec la loi $\mathcal{P}_{np}$, avec MatLab.

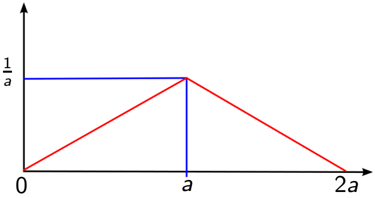
Applications de la loi de Poisson

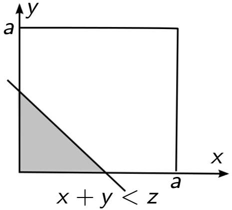
La loi de Poisson est utilisée dans des nombreuses applications techniques pour modéliser le nombre d'événements d'un certain type survenant pendant une période de temps fixée (par exemple, nombre de pannes par an, nombre d'arrivées dans une file d'attente, nombre de pièces défectueuses dans un lot de production etc.).

Loi uniforme continue (La loi fondamentale)

La loi uniforme sur $[0, a]$, notée $\mathcal{U}_{[0,a]}$ est définie par la densité uniforme $u_{[0,a]}(x) = \frac{1}{a}$ sur $[0, a]$, et 0 ailleurs. On a $E\{\mathcal{X}\} = \frac{a}{2}$, et $V\{\mathcal{X}\} = \frac{a^2}{12}$.

La somme de deux v.a. uniformes indépendantes est une loi triangulaire.





L'aire hachurée du graphique de droite illustre géométriquement la probabilité que la somme de deux variables aléatoires uniformes sur $[0, a]$ soit inférieure à une certaine valeur z. Si $\mathcal{Z} = \mathcal{X} + \mathcal{Y}$, on a que

$$F_{\mathcal{Z}}(z) = P(\mathcal{X} + \mathcal{Y} < z) = \frac{z^2}{2a^2}, \text{ lorsque } z \in [0, a] \text{ et } F_{\mathcal{Z}}(z) = 1 - \frac{(2a - z)^2}{2a^2}, \text{ lorsque } z \in [a, 2a].$$

On en déduit que la densité $f_{\mathcal{Z}}(z) = F'_{\mathcal{Z}}(z)$ vaut

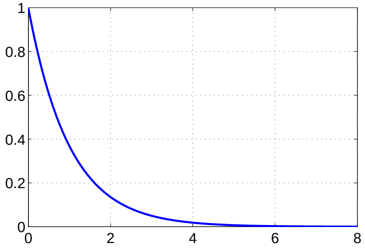
$$f_{\mathcal{Z}}(z) = \frac{z}{a^2}, \text{ lorsque } z \in [0, a] \text{ et } f_{\mathcal{Z}}(z) = \frac{(2a - z)}{a^2}, \text{ lorsque } z \in [a, 2a].$$

Loi exponentielle (Théoriquement très sympathique)

La densité de la loi (continue) exponentielle de paramètre λ est

$$f_{\mathcal{X}}(x) = \lambda \exp(-\lambda x) \text{ si } x > 0, \text{ 0 ailleurs.}$$

On a $E\{\mathcal{X}\} = \frac{1}{\lambda}$, et $V\{\mathcal{X}\} = \frac{1}{\lambda^2}$, et $F_{\mathcal{X}}(x) = 1 - \exp(-\lambda x)$.




```
x = 0:0.01:8;
lambda = 1;
y = lambda * exp(-lambda * x);
plot(x, y, 'LineWidth', 2);
```

Densité de probabilité de la loi exponentielle avec $\lambda = 1$, avec le code MATLAB permettant de générer la figure

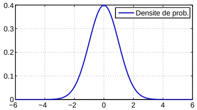
Loi Gaussienne ou normale (Très utilisée en pratique)

$\mathcal{X}$ suit une loi Gaussienne (ou normale), notée $\mathcal{N}(\mu, \sigma^2)$, si sa densité est

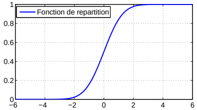
$$f_{\mathcal{X}}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right).$$

On a $E\{\mathcal{X}\} = \mu$, et $V\{\mathcal{X}\} = \sigma^2$.

Si $\mu = 0$ on dit que la loi est centrée. Si aussi $\sigma = 1$ on dit qu'elle est réduite.

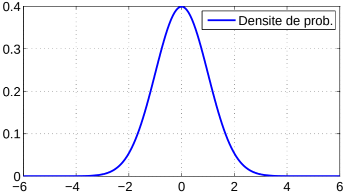



```
x = -6:0.01:6;
y = 1 / (sqrt(2 * pi)) * exp(-0.5 * x .^2);
plot(x, y, 'LineWidth', 2);
hold on;
```

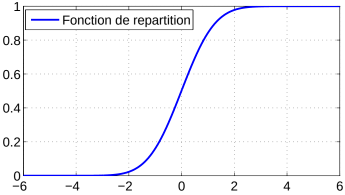



```
z = zeros(size(x));
for i=1:length(x)
    z(i) = sum(y(1:i));
end
z = z * (x(2) - x(1));
plot(x, z, 'LineWidth', 2);
```

Loi Gaussienne centrée réduite




```
x = -6:0.01:6;
y = 1 / (sqrt(2 * pi)) * exp(-0.5 * x .^2);
plot(x, y, 'LineWidth', 2);
hold on;
```




```
z = zeros(size(x));
for i=1:length(x)
    z(i) = sum(y(1:i));
end
z = z * (x(2) - x(1));
plot(x, z, 'LineWidth', 2);
```

$$f_{\mathcal{X}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \text{ et } F_{\mathcal{X}}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z.$$

Loi Gaussienne (propriétés principales)

Construction. Si $\mathcal{X}_1$ et $\mathcal{X}_2$ sont i.i.d $\mathcal{U}_{[0,1]}$, les variables (!!Erreur notes!!)

$$\mathcal{Y}_1 = \sqrt{-2\ln \mathcal{X}_1}\cos(2\pi\mathcal{X}_2) \quad (39)$$

$$\mathcal{Y}_2 = \sqrt{-2\ln \mathcal{X}_1}\sin(2\pi\mathcal{X}_2), \quad (40)$$

sont indépendantes et de loi normale centrée réduite $\mathcal{N}(0,1)$.

Transformation linéaire. Si $\mathcal{X} \sim \mathcal{N}(\mu,\sigma)$, alors $\mathcal{Z} = \alpha\mathcal{X} + \beta \sim \mathcal{N}(\alpha\mu+\beta, \alpha\sigma)$.

Simulation. Les deux propriétés précédents peuvent être utilisées pour simuler des v.a. Gaussiennes quelconques (cf. exercices MATLAB).

Additivité. Si $\mathcal{X} \sim \mathcal{N}(\mu_1,\sigma_1^2)$ et $\mathcal{Y} \sim \mathcal{N}(\mu_2,\sigma_2^2)$ sont **indépendantes**, leur somme suit encore une loi normale et on a

$$\mathcal{Z} = \mathcal{X} + \mathcal{Y} \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2).$$

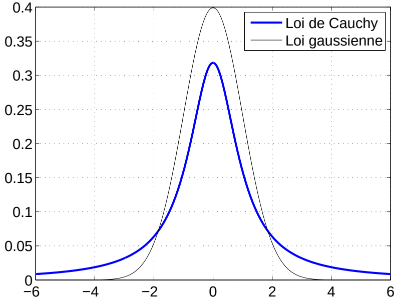
Par conséquent, la moyenne de n v.a. normales centrées réduites indépendantes est une variable normale centrée de variance $\frac{1}{\sqrt{n}}$.

Théorème central-limite. Il permet d'affirmer que la loi est d'application dans de nombreuses situations pratiques (voir plus loin).

Loi de Cauchy (Contre-exemple très important)

La densité de *Cauchy* est définie par $f_{\mathcal{X}}(x) = \frac{1}{\pi(1+x^2)}$.

Cette loi n'admet pas d'espérance (la densité décroit trop lentement pour que l'intégrale converge), ni a fortiori de variance.




```
x = -6:0.01:6;
g = 1 / (sqrt(2 * pi)) * exp(-0.5 *
c = 1 ./ (pi * (1 + x .^2));
plot(x, c, 'LineWidth', 2);
hold on;
plot(x, g, 'k');
```

Le rapport $\mathcal{Z} = \mathcal{X}_1/\mathcal{X}_2$ de deux v.a. i.i.d. $\mathcal{N}(0,1)$ suit une loi de Cauchy.

3.9 ○ ~~Convolution, fonctions caractéristiques, génératrices~~

Cette matière ne sera pas vue cette année, les pré-requis étant seulement enseignés en 3ème Bac.

Suggestion: Lire quand même la section 3.9 du syllabus lors de la révision du chapitre 3.

3.10 • Suites de v.a. et notions de convergence

Motivation de la notion de suite de variables aléatoires à valeurs réelles :

- La notion de suite de v.a. : on considère une famille $\{\mathcal{X}_n : n \in \mathbb{N}\}$ de variables aléatoires à valeurs réelles, définies sur un même espace de probabilité et indicées par $n \in \mathbb{N}$, et on s'intéresse aux propriétés de $\mathcal{X}_n$ lorsque $n \to \infty$.
- Applications : fondements de la statistique, de l'étude des processus stochastiques, de la théorie de l'information, de l'apprentissage statistique, et de l'algorithmique moderne.
- Outils techniques : différentes notions de convergence, utiles dans différents contextes.
- Débouchés immédiats : lois des grands nombres et théorème central limite, tous deux fondamentaux.

Convergence en probabilité (La plus utile en pratique)

Convergence en probabilité : Notation $(\mathcal{X}_n) \xrightarrow{P} a$

La suite $(\mathcal{X}_n)$ de v.a. réelles **converge en probabilité** vers la constante a, si $\forall \epsilon$ et η (arbitrairement petits), $\exists n_0$ tel que $n > n_0$ entraîne

$$P(|\mathcal{X}_n - a| > \epsilon) < \eta. \tag{41}$$

On dit que la suite de v.a. $(\mathcal{X}_n) \xrightarrow{P} \mathcal{X}$, si $(\mathcal{X}_n - \mathcal{X}) \xrightarrow{P} 0$.

Lancer de pièces n fois : Nous avons déjà démontré plus haut (grâce à l'inégalité de B-T) que $(\mathcal{Z}_n) \xrightarrow{P} 1/2$.

De façon plus générale : si $\mathcal{X}_n \sim \mathcal{B}(n, p)$ alors la suite $(\mathcal{Z}_n = \mathcal{X}_n/n) \xrightarrow{P} p$.

De façon encore plus générale : si $\mathcal{X}_n$ sont i.i.d et de variance finie, alors les v.a. $\overline{\mathcal{X}}_n \triangleq \left(\sum_{i=1} \mathcal{X}_n\right)/n$ forment une suite de v.a. telle que $(\overline{\mathcal{X}}_i) \xrightarrow{P} \mu$, avec $\mu = E\{\mathcal{X}_i\}, \forall i$. C'est aussi une conséquence directe de l'inégalité de B-T.

Convergence presque sûre (L'arme fatale)

Convergence presque sûre (ou forte) : Notation $(\mathcal{X}_n) \xrightarrow{p.s.} \mathcal{X}$

La suite $(\mathcal{X}_n)$ de v.a. réelles **converge presque sûrement** vers $\mathcal{X}$ si :

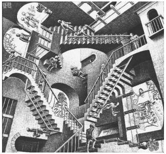
$$P(\{\omega | \lim_{n\to\infty} \mathcal{X}_n(\omega) \neq \mathcal{X}(\omega)\}) = 0. \tag{42}$$

Elle permet d'assimiler complètement $\mathcal{X}_n$ à $\mathcal{X}$ pour autant que n soit suffisamment grand. C'est pourquoi on l'appelle aussi convergence forte.

La convergence presque sûre implique la convergence en probabilité.

La démonstration du fait que la convergence 'p.s.' implique la convergence en probabilité est simple. Nous vous suggérons de la trouver vous même.

Convergence en moyenne d'ordre $p \in \mathbb{N}$



Convergence en moyenne d'ordre p : Notation $(\mathcal{X}_n) \xrightarrow{\mathcal{L}^p} \mathcal{X}$

Si $E\{(\mathcal{X}_n - \mathcal{X})^p\}$ existe $\forall n$, alors on a $(\mathcal{X}_n) \rightarrow \mathcal{X}$ **en moyenne d'ordre p** si $E\{|\mathcal{X}_n - \mathcal{X}|^p\} \rightarrow 0$.

Le cas pratique usuel est la moyenne quadratique ($p = 2$).

Cette notion sera étudiée plus en détails au chapitre 4, pour définir la géométrie des v.a. de carré intégrable.

La convergence en moyenne d'ordre p implique la convergence en probabilité.

Convergence en loi

Convergence en loi : Notation $(\mathcal{X}_n) \xrightarrow{\mathcal{L}} \mathcal{X}$

La suite $(\mathcal{X}_n)$ de v.a. réelles **converge en loi** vers $\mathcal{X}$ de fonction de répartition $F_{\mathcal{X}}(\cdot)$ si pour tout point de continuité x de $F_{\mathcal{X}}(\cdot)$, la suite $(F_{\mathcal{X}_n}(x))$ converge vers $F_{\mathcal{X}}(x)$. On note

$$(\mathcal{X}_n) \xrightarrow{\mathcal{L}} \mathcal{X}. \tag{43}$$

Il s'agit de la notion de convergence la plus faible.

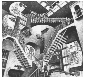
De fait, la convergence en probabilité implique la convergence en loi.

La convergence en loi est cependant très utilisée en pratique car elle permet d'approximer la fonction de répartition de $(\mathcal{X}_n)$ par celle de $\mathcal{X}$, et réciproquement.

On montre que si $F_{\mathcal{X}}(\cdot)$ est continue alors la convergence est uniforme (il s'agit d'un comportement plus régulier que la convergence ponctuelle, cf. cours d'analyse mathématique).

De plus, si les $F_{\mathcal{X}_n}(\cdot)$ admettent des densités alors la convergence en loi implique la convergence ponctuelle des densités.

Synthèse: relation entre ≠ notions de convergence



En moyenne d'ordre p

$\Rightarrow$

Presque sûre

$\Rightarrow$

En probabilité

$\Rightarrow$

En loi

3.11 Théorèmes de convergence (premier aperçu)

Convergence en loi 

Théorème de Moivre-Laplace: si $(\mathcal{X}_n)$ forme une suite de v.a. binomiales $\mathcal{B}(n,p)$, alors

$$\left(\frac{\mathcal{X}_n - np}{\sqrt{np(1-p)}}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \quad (44)$$

Utile en pratique pour approximer une loi binomiale par une loi Gaussienne.

Théorème central-limite: si $(\mathcal{X}_n)$ forme une suite de v.a. i.i.d. de moyenne μ et d'écart-type σ (ces deux moments sont donc supposés exister), alors

$$\left(\frac{\sum_{i=1}^{n} \mathcal{X}_i - n\mu}{\sigma\sqrt{n}}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \quad (45)$$

Ce théorème établit la convergence en loi vers la loi normale d'une somme de v.a. "i.i.d." sous des hypothèses très peu contraignantes. On retrouve comme cas particulier le théorème de Moivre-Laplace, en prenant des variables de Bernoulli.

Contre-exemple : loi de Cauchy.

Loi faible des grands nombres

Soient $\mathcal{X}_i, \forall i = 1, \ldots, n$ indépendantes et d'espérances μ_i finies et de variances σ_i finies, et telles que $\frac{1}{n}\sum_{i=1}^{n} \mu_i \to \mu$ et $\frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 \to 0$, alors $\overline{\mathcal{X}}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i$ est telle que

$$(\overline{\mathcal{X}}_n) \xrightarrow{P} \mu. \tag{46}$$

Cas particulier : les v.a. $\mathcal{X}_i$ sont i.i.d. μ, σ. On a alors $\frac{1}{n}\sum_{i=1}^{n} \mu_i = \mu$ et $\frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = \frac{\sigma^2}{n}$.

Lois forte des grands nombres 

Soient $\mathcal{X}_i, \forall i = 1, \ldots, n$ indépendantes et d'espérances μ_i finies, et telles que $\frac{1}{n}\sum_{i=1}^{n} \mu_i \to \mu$ et de variances σ_i finies, telles que $\sum_{i=1}^{n} \frac{\sigma_i^2}{i^2} \to a$, alors la suite $\overline{\mathcal{X}}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i$ est telle que

$$(\overline{\mathcal{X}}_n) \xrightarrow{p.s.} \mu. \tag{47}$$

Cas particulier : les v.a. $\mathcal{X}_i$ sont i.i.d. μ, σ. On a alors $\frac{1}{n}\sum_{i=1}^{n} \mu_i = \mu$ et $\sum_{i=1}^{n} \frac{{\sigma_i}^2}{i^2} = \sigma^2 \sum_{i=1}^{n} \frac{1}{i^2}$, qui converge.

Notez bien que $\left(\sum_{i=1}^{n} \frac{\sigma_i^2}{i^2} \to a\right) \Rightarrow \left(\sum_{i=1}^{n} \frac{\sigma_i^2}{n^2} \to 0\right)$, la réciproque n'étant pas vraie.

La loi forte implique la convergence 'p.s' qui implique elle-même la convergence en probabilité de la loi faible!

3.12 Problèmes et applications

- Résolution de problèmes pratiques à l'aide des notions développées dans le chapitre 3.
  - Evaluation de la fiabilité d'un système.
  - Evaluation du coût d'exploitation d'un système
  - Lire la section 3.12.1
- Méthode de Monte-Carlo
  - Lire la section 3.12.2
  - Faire le travail 2

Monte-Carlo : pour estimer une intégrale simple

- Le calcul de
$$I_g = \int_0^1 g(x)\,\mathrm{d}x. \quad (48)$$
est équivalent à la détermination de l'espérance mathématique d'une variable aléatoire $\mathcal{Y} = g(\mathcal{X})$ avec $\mathcal{X} \sim \mathcal{U}[0,1]$.
- Si nous supposons que $\sigma_{\mathcal{Y}}$ est finie (p.ex. si $g(x)$ est continue), et que nous disposons d'un échantillon i.i.d. de valeurs y_i de $\mathcal{Y}$, nous pouvons estimer cette espérance par
$$\hat{I}_g = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad (49)$$
l'écart-type de l'estimateur étant donné par
$$\sigma_{\hat{I}_g} = \frac{1}{\sqrt{n}}\sigma_{\mathcal{Y}}. \quad (50)$$
- En pratique, on peut générer un échantillon de valeurs de $\mathcal{Y}$ en utilisant un générateur de nombres aléatoires $x_i \sim \mathcal{U}[0,1]$, et en calculant $y_i = g(x_i)$. Cf. Travail no 2.

Monte-Carlo : pour estimer une intégrale multiple

- Le calcul de
$$I_g = \int_{[0,1]^p} g(x_1, \ldots, x_p)\, \mathrm{d}x_1 \cdots \mathrm{d}x_p. \tag{51}$$
est équivalent à la détermination de l'espérance mathématique d'une variable aléatoire $\mathcal{Y} = g(\mathcal{X}_1, \ldots, \mathcal{X}_p)$ avec $\mathcal{X}_i$ i.i.d. $\sim \mathcal{U}[0, 1]$.
- Si nous supposons que $\sigma_{\mathcal{Y}}$ est finie (p.ex. si $g(x_1, \ldots, x_p)$ est continue), et que nous disposons d'un échantillon i.i.d. de valeurs y_i de $\mathcal{Y}$, nous pouvons estimer cette espérance par
$$\hat{I}_g = \frac{1}{n} \sum_{i=1}^{n} y_i, \tag{52}$$
l'écart-type de l'estimateur étant donné par
$$\sigma_{\hat{I}_g} = \frac{1}{\sqrt{n}} \sigma_{\mathcal{Y}}. \tag{53}$$
- En pratique, on peut générer un échantillon de valeurs de $\mathcal{Y}$ en utilisant un générateur de nombres aléatoires $x_{i,j} \sim \mathcal{U}[0, 1]$, et en calculant $y_i = g(x_{i,1}, \ldots, x_{i,p})$. Cf. Travail no 2.

Monte-Carlo : réduction de la variance (a)

- Echantillonnage d'importance
  - Soit $\hat{I}_h^{f_\mathcal{X}} = \frac{1}{n}\sum_{i=1}^n h(x_i)$ calculé en moyennant les valeurs de la fonction $h(\cdot)$ appliquée aux valeurs x_i supposées i.i.d. $\sim f_\mathcal{X}$.
  - On a $E\{\hat{I}_h^{f_\mathcal{X}}\} = E\{h(\mathcal{X})\}$ et $V\{\hat{I}_h^{f_\mathcal{X}}\} = V\{h(\mathcal{X})\}/n$, avec
$$E\{h(\mathcal{X})\} = \int_{\Omega_\mathcal{X}} h(x) f_\mathcal{X}(x)\,\mathrm{d}x.$$
  - En prenant $h(x) = \frac{g(x)}{f_\mathcal{X}(x)}$ nous obtenons un nouvel estimateur de l'intégrale $I_g = \int_{\Omega_\mathcal{X}} g(x)\,\mathrm{d}x$, donné par
$$\hat{I}_{g/f_\mathcal{X}}^{f_\mathcal{X}} = \frac{1}{n}\sum_{i=1}^n \frac{g(x_i)}{f_\mathcal{X}(x_i)},$$
dont la variance sera d'autant plus faible que $h(x) = \frac{g(x)}{f_\mathcal{X}(x)}$ est proche d'une constante.
  - Idéalement, on prendrait $f_\mathcal{X}^*(x) = g(x)/I_g$, auquel cas la variance de l'estimateur serait nulle !

Monte-Carlo : réduction de la variance (b)

- Variable de contrôle
  - Soit une fonction $h(x)$ dont nous connaissons déjà l'intégrale I_h avec grande précision.
  - Nous pouvons alors calculer I_g par

$$I_g = I_h + \int_{\Omega_{\mathcal{X}}} (g(x) - h(x))\,\mathrm{d}x.$$

  - Si $g(x) - h(x)$ varie peu par rapport à h sur l'intervalle d'intégration, on pourra appliquer la méthode de Monte-Carlo à $g - h$, ce qui nécessitera un nombre plus faible d'échantillons, à précision égale, mais une double évaluation de fonction pour chaque échantillon.
  - Idéalement, on prendrait $h(x) = g(x)$ auquel cas la variance de l'estimateur serait nulle !

Monte-Carlo : réduction de la variance (c)

- Echantillonnage stratifié
  - Soit un événement A, de probabilité ϵ-proche de 0.5 dont nous souhaitons connaître de façon très précise la probabilité.
  - Nous pouvons évaluer l'espérance de 1_A par Monte-Carlo. Pour n réalisations de l'expérience, nous aurons $\hat{P}(A) = n^{-1} \sum_{i=1}^{n} 1_A(\omega_i)$, estimateur dont l'erreur standard $\sigma_{\hat{P}}$ sera proche de $0.5/\sqrt{n}$.
  - Supposons que l'événement B soit tel que $P(A|B)$ est ϵ-proche de 1 et que $P(A|B^c)$ est ϵ-proche de 0, que nous connaissions $P(B)$ (et donc $P(B^c)$), et que nous puissions faire deux expériences séparées respectivement selon $P(A|B)$ et selon $P(A|B^c)$.
  - En allouant $n/2$ observations à B (respectivement à B^c), nous pouvons estimer $P(A|B)$ avec une erreur-standard proche de 0, de même que $P(A|B^c)$, et puis calculer $\hat{P}(A) = P(B)\hat{P}(A|B) + P(B^c)\hat{P}(A|B^c)$.
  - L'erreur-standard de cet estimateur combiné sera donc proche de 0.
  - Idéalement, on prendrait B tel que $P(A|B) = 1$ et $P(A|B^c) = 0$, auquel cas la variance de l'estimateur combiné serait nulle !

Méthode de Monte-Carlo : discussion

- Les lois des grands nombres, appliquées aux variables i.i.d., montrent que le calcul de l'espérance d'une v.a. peut (souvent, mais pas toujours) être réalisé en calculant une moyenne des valeurs observées lors d'un grand nombre de réalisations d'expériences répétées.
- La méthode de Monte-Carlo consiste à utiliser l'ordinateur pour faire ce genre d'expérience. On exploite un générateur de nombres aléatoires pour simuler l'observation d'une expérience aléatoire, et on répète le processus un grand nombre de fois, en calculant la valeur moyenne des observations.
- Cette technique est illustrée dans le cadre du travail pratique numéro 2 de ce cours.
- La méthode de Monte-Carlo et les techniques de réduction de la variance qui lui sont associées sont à la base de nombreux progrès en ingénierie et en informatique.