LBIRE 1203
Probabilités

P. Bogaert

Université catholique de Louvain
UCL/AGRO/ELIE
Croix du Sud 2/16
1348 Louvain-la-Neuve, Belgique

patrick.bogaert@uclouvain.be

Informations

Contacts

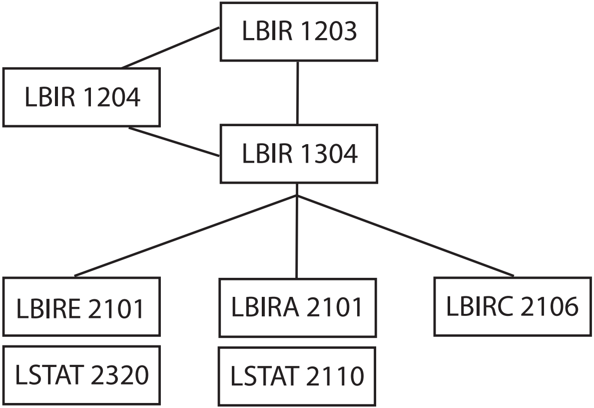
- Titulaire : Patrick Bogaert (ENGE)
  Bâtiment Mendel, Local C 313
  Tél : 010/47.36.82
  E-mail : patrick.bogaert@uclouvain.be

- Assistant : Sarah Gengler (ENGE)
  Bâtiment Mendel, Local C 305
  Tél : 010/47.36.11
  E-mail : sarah.gengler@uclouvain.be

Enseignement

- **Support :** Livre *Probabilités pour scientifiques & ingénieurs*
  Cahier d'exercices
- **Travaux :** Deux à trois travaux à préparer
  A remettre à date fixée
- **Examen :** A livre ouvert
  Uniquement à l'aide du support original
  Ne comprend que des exercices
- **Conseils :** Lire le support
  Ne rien mémoriser, tout comprendre
  Etre à jour dans la matière
  Préparer les T.P.

Structure du cursus "Probabilités-Statistique"



Acquis d'apprentissage

A la fin de cette activité, l'étudiant est capable de :

- Nommer, décrire et expliquer les concepts théoriques relatifs à la théorie des probabilités ;
- Manipuler les expressions mathématiques de manière formelle et avec une notation rigoureuse en vue d'en déduire de nouvelles expressions utiles ou des résultats théoriques recherchés ;
- Reformuler l'énoncé textuel d'un problème dans un formalisme mathématique et probabiliste non ambigu, en utilisant les concepts et outils théoriques adéquats ;
- Résoudre un problème appliqué en suivant une approche déductive basée sur la manipulation correcte et utile des expressions ;
- Valider la cohérence interne de la formalisation et de la solution d'un problème de calcul des probabilités sur base des contraintes logiques induites par la théorie.

Introduction

Différence entre Probabilités et Statistique

- Probabilités : Modèle mathématique pour analyser les phénomènes aléatoires (où le hasard intervient).

- Statistique : Apprendre en observant les phénomènes aléatoires. Nécessité d'estimer, d'inférer.

"En Probabilités, Dieu nous donne les paramètres et nous déterminons ce qui va s'ensuivre. En Statistique, les choses se sont déjà produites et nous essayons de déterminer les paramètres que Dieu a choisis."

Exemples – probabilités

- On sait que la probabilité d'obtenir pile est de 0.5.
  - quelle est la probabilité d'obtenir deux fois pile sur 2 jets ?
  - quelle est la probabilité de ne pas obtenir pile sur 10 jets ?
- On sait que parmi 100 appareils, 3 sont défectueux.
  - quelle est la probabilité qu'il n'y ait aucun appareil qui soit défectueux parmi 10 appareils pris au hasard ?
  - quelle est la taille maximale d'un lot aléatoire telle qu'il ne contient aucun appareil défectueux avec une probabilité au moins égale à 90 % ?

Exemples – statistique

- On observe la production de 10 vaches de race A et de 10 vaches de race B.
  - ☐ quelle est la production moyenne de chaque race ?
  - ☐ quelle est la variabilité de la production de chaque race ?
  - ☐ peut-on admettre que les races ont la même production ?
- On observe des colonies de bactéries sur 100 plaques de Pétri.
  - ☐ quelle est le nombre moyen de colonies par plaque ?
  - ☐ quelle est la probabilité que j'observe 1,2,3,... colonies ?

Résumé

Les **Probabilités** font appel à la notion de **modèle**. Quels modèles et hypothèses va-t-on utiliser pour effectuer ces calculs ?

La **Statistique** fait appel à des observations et **estime** ou **vérifie** la pertinence d'un modèle/de ses paramètres.

La maîtrise du calcul des Probabilités est un pré-requis essentiel pour aborder la Statistique :

Pour pouvoir estimer la pertinence d'un modèle à partir d'observations, il faut pouvoir déterminer ce qu'on s'attend à observer sur base de ce modèle.

Illustration : jet d'un dé

- Si l'on suppose un dé équilibré et les jets successifs indépendants, on peut calculer toutes les probabilités désirées. Par exemple :
  - ☐ la probabilité d'obtenir au moins un "6" en 3 jets
  - ☐ la probabilité d'obtenir un "6" pour les 3 jets

- Pour tester le fait que le dé est équilibré, on effectue une longue série de jets et on observe la fréquence de chaque résultat. Comment décider si l'hypothèse est acceptable ou non lorsque :
  - ☐ sur 5 jets, on obtient le même résultat ?
  - ☐ sur 50 jets, on obtient le même résultat ?

Pour répondre à cette question, on a besoin des Probabilités.

1. Notions de base

Epreuve ou expérience aléatoire

- expérience où le hasard intervient (résultat inconnu à l'avance)
- l'ensemble des résultats possibles Ω peut être décrit
- l'expérience est répétable sous des conditions semblables

On doit définir $\Omega = \{\omega : \omega \text{ est un résultat possible}\}$.
L'ensemble Ω est aussi appelé l'*univers* de l'expérience.
Cet ensemble Ω peut contenir :

- ☐ un nombre fini d'éléments
- ☐ une infinité dénombrable d'éléments
- ☐ une infinité non dénombrable d'éléments

Exemples

- nombre fini d'éléments :

  (a) jet d'une pièce : $\Omega = \{P, F\}$
  (b) jet successif de deux pièces : $\Omega = \{PP, FF, PF, FP\}$
  (c) jet d'un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$
  (d) jet de deux dés :

  $$\Omega = \{\{1,1\}, \ldots, \{1,6\}, \{2,1\}, \ldots, \{2,6\}, \ldots, \{6,1\}, \ldots, \{6,6\}\}$$
  $$\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

- infinité dénombrable d'éléments :

  (a) nombre de colonies sur une plaque de Pétri : $\Omega = \{0, 1, 2, \ldots\}$
  (b) âge d'un personne majeure prise au hasard : $\Omega = \{18, 19, \ldots\}$

Exemples (suite)

- infinité non dénombrable d'éléments :

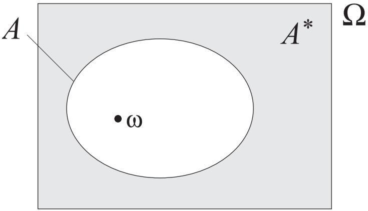
  (a) volume de lait produit par une vache :$\Omega = \{\omega : \omega \geq 0\}$
  (b) température au 1er janvier : $\Omega = \{\omega : \omega \in [-273^o, \infty)\}$

Remarque : l'ensemble Ω peut être défini :

- ☐ en extension (on énumère un à un les éléments)
- ☐ en compréhension (on définit ces éléments par une propriété)

Evénement

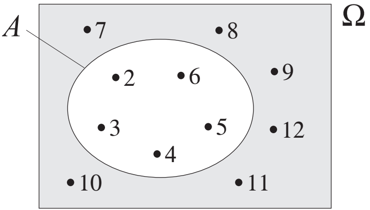
- sous-ensemble A de Ω possédant une propriété
- on parle d'évènement réalisé (A) ou non réalisé (A^*)



A est réalisé si le résultat ω de l'épreuve $\in A$
A n'est pas réalisé si le résultat ω de l'épreuve $\notin A$

Exemples

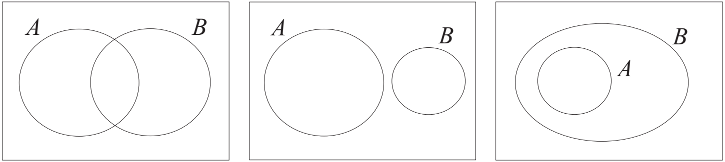
(a) jet d'une pièce : $A = \{P\}$
(b) jet successif de deux pièces : $A = \{PP, FF\}$
(c) jet de deux dés : $A =$"la somme est inférieure à 7"



- ω : évènement élémentaire (résultat de l'épreuve)
- Ω : évènement certain (tous les résultats possibles)
- $\emptyset$: évènement impossible (ensemble vide)

Relations entre évènements

- si deux évènements A et B ont certains éléments en commun, alors $A \cap B$ est l'ensemble de ces éléments
- si deux évènements A et B n'ont pas d'éléments communs, il sont incompatibles : $A \cap B = \emptyset$
- si un évènement A implique un évènement B, alors $A \subset B$



Composition d'évènements

- négation ($*$) d'un évènement :

  A^* est réalisé si A n'est pas réalisé
  A^* est le complémentaire de A par rapport à Ω

- union ($\cup$) de deux évènements A et B :

  $A \cup B$ est réalisé si A *ou* B est réalisé

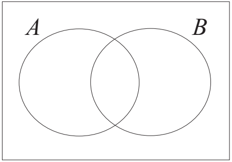
- intersection ($\cap$) de deux évènements A et B :

  $A \cap B$ est réalisé si A et B sont réalisés

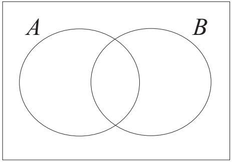
Exercice : Représentez les évènements sur les graphiques

(a) $A \cup B$ (b) $A \cup B^*$ (c) $A^* \cup B^*$
(d) $A \cap B$ (e) $A^* \cap B$ (f) $A^* \cap B^*$

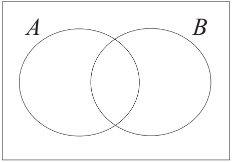
(a)



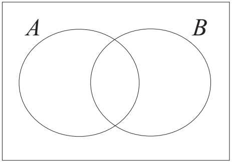
(b)



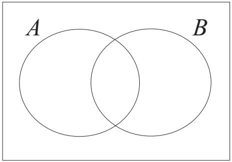
(c)



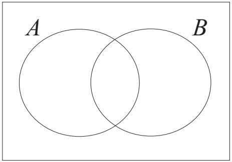
(d)



(e)



(f)



Algèbre des évènements

- idempotence: $A \cup A = A \qquad A \cap A = A$
- commutativité : $A \cup B = B \cup A \qquad A \cap B = B \cap A$
- associativité : $$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$$ $$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$$
- distributivité : $$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$ $$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$
- identité : $$A \cup \emptyset = A \quad A \cup \Omega = \Omega$$ $$A \cap \emptyset = \emptyset \quad A \cap \Omega = A$$

Exercice : vérifiez graphiquement toutes ces égalités

Algèbre des évènements

- complémentarité : $$\begin{array}{lll} A \cup A^* = \Omega & A \cap A^* = \emptyset & \\ (A^*)^* = A & \emptyset^* = \Omega & \Omega^* = \emptyset \end{array}$$

- lois de *de Morgan* : $(A \cup B)^* = A^* \cap B^*$ $\quad (A \cap B)^* = A^* \cup B^*$

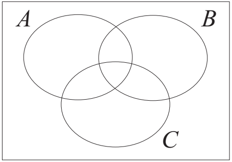
Les lois de *de Morgan* sont particulièrement utiles pour simplifier certaines expressions (raisonnement par le contraire).

Exercice : Représentez les évènements sur les graphiques

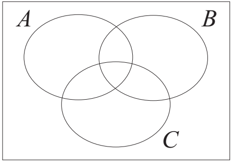
(a) $A^* \cap B^* \cap C^*$ (b) $A^* \cup B^* \cup C^*$ (c) $(A \cup B) \cap C$

(d) $(A^* \cap B^*) \cup C^*$ (e) $(A \cap B) \cup C$ (f) $(A^* \cup B^*) \cap C^*$

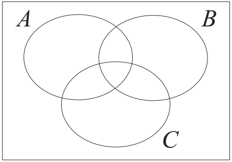
(a)



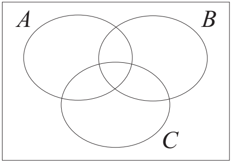
(b)



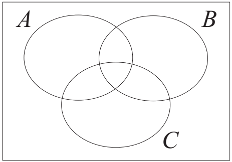
(c)



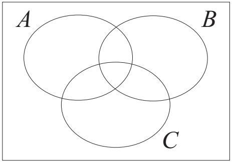
(d)



(e)



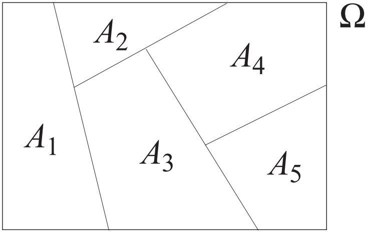
(f)



Système complet d'événements

Une famille d'évènements $\{A_1, \ldots, A_i, \ldots, A_n\}$ forme un système complet d'évènements (une partition de Ω) ssi :

- aucun des évènements n'est impossible : $\forall i, A_i \neq \emptyset$
- les évènements sont incompatibles deux à deux : $\forall i \neq j, A_i \cap A_j = \emptyset$
- l'union des évènements est l'évènement certain : $\bigcup_{i=1}^{n} A_i = \Omega$



Exercice

Les évènements suivants forment-il un système complet ?

- $A = \{1, 3, 5\}$ et $B = \{2, 4, 6\}$ pour le jet d'un dé.
- $A_1 = \{\omega : \omega \geq 0\}$, $A_2 = \{\omega : \omega \geq 5\}$, $A_3 = \{\omega : \omega \geq 10\}$ pour le nombre ω de colonies sur une plaque de Pétri.
- $A_1 = \{1\}$, $A_2 = \{2\}$, $A_3 = \{3\}$ pour le nombre de pièces qui sont défectueuses dans un lot de 3 pièces.

2. Notions de probabilités

Définitions historiques

Comment définir la probabilité $P(A)$ d'un événement A à partir d'une expérience aléatoire ?

- *Définition classique* (Laplace, 1819)

$$\begin{array}{ll} \#A & \text{nombre de cas équiprobables réalisant } A \\ \#\Omega & \text{nombre de cas équiprobables possibles} \end{array}$$

$$\Longrightarrow P(A) = \frac{\#A}{\#\Omega}$$

- Problèmes :
  - les cas doivent être équiprobables
  - le nombre de cas possible doit être fini
  - la définition dépend du terme "équiprobable"

Exemple : obtenir 7 avec deux dés ➡

Si on s'intéresse à la somme des points obtenus, on a

$$\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\} \quad \Longrightarrow \quad \#\Omega = 11$$

L'événement $A \equiv$"la somme vaut 7" représente un cas parmi les 11.
On en déduirait donc que

$$P(A) = \frac{\#A}{\#\Omega} = \frac{1}{11}$$

Ce résultat est bien entendu incorrect.

Question : quelle est la valeur correcte de $P(A)$?

Définitions historiques

- *Définition en fréquence*

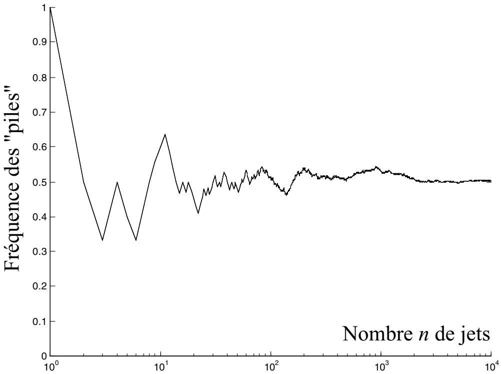
n	nombre de répétitions de l'expérience
$n(A)$	nombre de réalisations de A parmi les n répétitions

$$\Longrightarrow P(A) = \lim_{n\to\infty} \frac{n(A)}{n}$$

- Problèmes : – probabilité définie comme une limite
  – définition peu commode à utiliser

Exemple : fréquence relative des "piles" pour une pièce

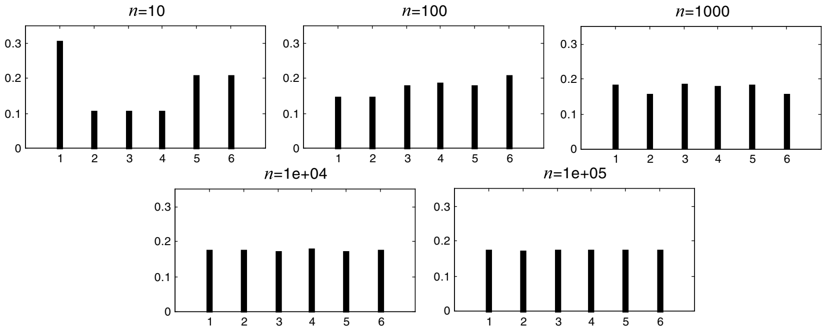
On peut tracer l'évolution de la fréquence relative en fonction de n :



La fréquence relative tend vers la valeur théorique de 1/2.

Exemple : Fréquences relatives des résultats pour un dé

On peut tracer l'évolution de la fréquence relative de chacun des 6 résultats en fonction de n :



Chaque fréquence relative tend vers la valeur théorique de 1/6.

Définition axiomatique (Kolmogorov, 1933)

Axiome = *proposition que l'on choisit de considérer comme vraie, sans démonstration, et qui sert de point de départ pour les développements théoriques.*

- Axiome 1 : la probabilité est un nombre non négatif
  $P(A) \geq 0$

- Axiome 2 : la probabilité de l'événement certain vaut 1
  $P(\Omega) = 1$

- Axiome 3 : pour deux événements incompatibles, la probabilité de leur union vaut la somme de leur probabilité
  $P(A \cup B) = P(A) + P(B)$ si $A \cap B = \emptyset$

Définitions

La définition axiomatique permet de dériver aisément toutes les formules nécessaires **uniquement** sur base des 3 axiomes.

Par exemple :

- $P() = 0$
- $P(A) \leq 1$
- $P(A^*) = 1 - P(A)$
- $P(A) \leq P(B) \quad$ si $A \subseteq B$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Exemple

Messieurs A et B sont des tireurs à l'arc.

- la probabilité que A touche la cible vaut 0.7
- la probabilité que B touche la cible vaut 0.5
- si A et B tirent en même temps, la probabilité qu'ils touchent tous les deux la cible vaut 0.35

Quelle est la probabilité qu'au moins une flèche touche la cible si A et B tirent en même temps ?

Définitions

En utilisant les lois de l'algèbre, on peut généraliser les résultats à plus de deux événements :

- $$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

- $$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \quad \text{si } A_i \cap A_j = \emptyset \quad \forall i \neq j$$

Exemple

On s'intéresse au poids W d'individus masculins

- $A \equiv$"un individu pèse au moins 55 kg"
- $B \equiv$"un individu pèse au plus 105 kg"
- $C \equiv$"un individu pèse au moins 55 kg et au plus 105 kg"

On donne $P(A) = 0.95$ et $P(B) = 0.97$
Que vaut $P(C)$?

Indice : Définissez les événements supplémentaires suivants :

- $A^* \equiv$"un individu pèse moins de 55 kg"
- $B^* \equiv$"un individu pèse plus de 105 kg"

Probabilité conditionnelle

Pour évaluer la fréquence d'une maladie bovine et l'efficacité d'un test sanguin, on fait des comptages dans un troupeau de n vaches.

$M \equiv$"la vache est malade" $\quad T \equiv$"le test est positif"
$M^* \equiv$"la vache est saine" $\quad T^* \equiv$"le test est négatif"

	M	M^*	Total
T	$n(M \cap T)$	$n(M^* \cap T)$	$n(T)$
T^*	$n(M \cap T^*)$	$n(M^* \cap T^*)$	$n(T^*)$
Total	$n(M)$	$n(M^*)$	n

Quelle est :
- la fréquence de M quand T est réalisé ?
- la fréquence de T^* quand M est réalisé ?

Probabilité conditionnelle

- Définition de la probabilité conditionnelle $P(A|B)$:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{avec } P(B) \neq 0$$

- Elle vérifie les 3 axiomes de base :
  - $P(A|B) \geq 0$
  - $P(\Omega|B) = 1$
  - $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$ si $A_1 \cap A_2 = \emptyset$

Probabilité conditionnelle

- Les probabilités conditionnelles ont des propriétés semblables aux probabilités non conditionnelles :

  - $P(|B) = 0$
  - $P(A|B) \leq 1$
  - $P(A^*|B) = 1 - P(A|B)$
  - $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$

Exercice

Dans un troupeau de 100 vaches, 7 sont malades. On sait qu'un test est positif dans 98 % des cas quand la vache est malade. On prend une vache au hasard. Quelle est la probabilité qu'elle soit malade et que le test soit positif ?

Théorème des probabilités composées

On sait que : $P(A|B) = \dfrac{P(A \cap B)}{P(B)} \qquad P(B|A) = \dfrac{P(A \cap B)}{P(A)}$

On a donc : $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

On peut généraliser pour 3 événements :

$$\begin{aligned} P(A \cap B \cap C) &= P(A|(B \cap C))P(B \cap C) \\ &= P(A|(B \cap C))P(B|C)P(C) \end{aligned}$$

L'intersection d'événements peut s'exprimer comme un produit de probabilités conditionnelles

Théorème des probabilités composées

De manière générale, pour n événements :

$$
\begin{aligned}
& P(A_1 \cap \cdots \cap A_n) \\
& \\
& = P(A_1 \cap \cdots \cap A_{n-1}) P(A_n | (A_1 \cap \cdots \cap A_{n-1})) \\
& \vdots \\
& = P(A_1) P(A_2 | A_1) P(A_3 | (A_1 \cap A_2)) \cdots P(A_n | (A_1 \cap \cdots \cap A_{n-1}))
\end{aligned}
$$

Exercice

Dans le troupeau de 100 vaches dont 7 sont malades, on sélectionne n vaches au hasard, qu'on sépare du troupeau. On veut calculer la probabilité qu'aucune des vaches sélectionnées ne soit malade quand $n = 1, 2, 3, \ldots$

Théorème des probabilités totales

Soient des événements $A_1, A_2, \ldots, A_n$ formant une partition.
Soit un événement B quelconque.

On montre alors que :

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$

Preuve :

Exercice

Un lot de 21 pièces est issu de trois techniques de fabrication différentes.

Technique n^o	1	2	3
Probabilité de défectuosité	0.01	0.02	0.005
Nombre de pièces produites	7	9	5

Quelle est la probabilité qu'une pièce prise au hasard soit défectueuse ?

Théorème de Bayes

Soient des événements $A_1, A_2, \ldots, A_n$ formant une partition.
Soit un événement B quelconque.

On montre alors que :

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)} \qquad \forall i = 1, \ldots, n$$

Preuve :

Exercice

Un lot de 21 pièces est issu de trois techniques de fabrication différentes.

Technique n^o	1	2	3
Probabilité de défectuosité	0.01	0.02	0.005
Nombre de pièces produites	7	9	5

On prend une pièce au hasard. Sachant qu'elle est défectueuse, quelle est la probabilité qu'elle soit issue de la technique n^o i ?

Théorème de Bayes

- Le théorème de Bayes est aussi appelé le *théorème des causes* :

  On connaît la probabilité $P(B|A_i)$ d'un événement B pour chacune des causes $A_1, \ldots, A_n$ possibles. Si l'on observe que B s'est réalisé, quelle est la probabilité $P(A_i|B)$ que ce soit dû à la cause A_i (avec $i = 1, \ldots, n$) ?

- *Exemples d'applications*

  - systèmes experts
  - filtres bayésien anti-spam (e-mail)

Exercice

On place dans un étang des poissons rouge et gris. Un tiers de ces poissons sont d'origine chinoise. Parmi les poissons chinois, 30 % sont rouges. Au total, 20 % des poissons sont rouges. Un promeneur voit un poisson rouge dans l'étang. Quelle est la probabilité qu'il soit d'origine chinoise ?

Exercice ➡

Dans une université, 10 % des étudiants sont d'origine chinoise. Parmi ceux-là, 90 % ont moins de 30 ans. Il y a 5 % d'étudiants chinois ayant moins de 30 ans et effectuant un doctorat. Vous rencontrez par hasard un étudiant qui dit être chinois et avoir moins de 30 ans. Quelle est votre chance de gagner en pariant qu'il effectue un doctorat ?

Exercice

Deux traitements peuvent être appliqués à une plante malade, avec des probabilités de guérison égales à 90 % et 80 %. Un tiers des plantes reçoivent le premier traitement, les autres reçoivent le second traitement. Si une plante malade est guérie, quelle est la probabilité que le premier traitement ait été appliqué ?

Indépendance

- L'indépendance entre A et B est définie par la propriété :

$$A \perp B \quad \Longleftrightarrow \quad P(A \cap B) = P(A)P(B)$$

L'événement A est indépendant de l'événement B si la (non) réalisation de B n'apporte aucune information sur la probabilité de réalisation de A, et vice-versa :

$$P(A|B) = P(A|B^*) = P(A) \qquad P(B|A) = P(B|A^*) = P(B)$$

- Il s'agit d'une notion fondamentale très importante.

Exercice

On teste la présence d'une maladie dans un troupeau de 100 vaches dont 7 sont malades. Deux possibilités existent quand une vache a été testée:

- ☐ elle est retirée du troupeau et ne sera plus testée
- ☐ elle est replacée dans le troupeau sans marque distinctive

Dans les deux cas, quelle est la probabilité qu'aucune des 7 vaches malades ne soit testée si l'on choisit au hasard 10 vaches parmi les 100 vaches du troupeau ?

Remarques

- ne pas confondre indépendance et incompatibilité !
  - ☐ indépendance : $A \perp B \implies P(A \cap B) = P(A)P(B)$
  - ☐ incompatibilité : $A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$
- l'indépendance peut être généralisée à plus de 2 événements:

$$A \perp B \perp C \iff \begin{cases} P(A \cap B \cap C) = P(A)P(B)P(C) \\ P(A \cap C) = P(A)P(C) \\ P(B \cap C) = P(B)P(C) \\ P(A \cap B) = P(A)P(B) \end{cases}$$

- poser l'hypothèse d'indépendance a d'importantes conséquences.

3. Variables aléatoires

Définition

- C'est une relation entre des événements associés à une expérience aléatoire et des valeurs numériques.

  Soit A un événement défini sur l'ensemble Ω.
  Soit B un événement défini sur un ensemble $R_X \subseteq \mathbb{R}$.

  La variable aléatoire X est une relation de $\Omega \to R_X$ telle que $\forall B \subseteq R_X$,

$$A = \{\omega : X(\omega) \in B\} \quad \text{avec } P(A) = P(B)$$

- Les valeurs R_X constituent le *domaine de variation*.

Exemple : jet d'une pièce

- Soient les événements suivants :

$$A \equiv \text{“Pile”}; \quad A^* \equiv \text{“Face”}; \quad \Omega = \{A, A^*\}$$

- Définissons les événements équivalents suivants :

$$B \equiv (X = 1); \quad B^* \equiv (X = 0)$$

On a ainsi défini une variable aléatoire X avec $R_X = \{0, 1\}$.

Exemple : jeu de dé

- Soit le jeu où, en jetant un dé, on gagne
  - ☐ 2 Euros si le résultat est “1”,“2” ou “3”
  - ☐ 3 Euros si le résultat est “4” ou “5”
  - ☐ 6 Euros si le résultat est “6”

  On a donc $\Omega = \{“1”,“2”,“3”,“4”,“5”,“6”\}$.

- Des événements équivalents sont donnés par :

$$
\begin{aligned}
(X = 2) &\equiv \{“1”,“2”,“3”\} \\
(X = 3) &\equiv \{“4”,“5”\} \\
(X = 6) &\equiv \{“6”\}
\end{aligned}
$$

  On a ainsi défini une variable aléatoire X avec $R_X = \{2, 3, 6\}$.

Variable aléatoire discrète

- Le nombre de valeurs possibles pour X est fini ou est une infinité dénombrable.

Exemples :

- nombre de vaches malades dans un troupeau de n vaches
  $R_X =$
- nombre d'insectes par plant de tabac
  $R_X =$
- nombre d'enfants dans une famille nombreuse
  $R_X =$

Variable aléatoire discrète

- On peut caractériser une variable aléatoire discrète par :
  - sa fonction de probabilité $p(x)$
  - sa fonction de répartition $F(x)$

- Les valeurs de ces fonctions peuvent être données par :
  - une formule mathématique
  - une table des probabilités $P(X = x)$ ou $P(X \leq x)$
  - des graphiques

Le choix dépend essentiellement du type de problème rencontré.

Fonction de probabilité

- Fonction donnant $\forall x \in \mathbb{R}$ la probabilité $P(X = x)$ correspondante. On note en abrégé

$$p(x) = \begin{cases} P(X = x) & \text{pour } x \in R_X \\ 0 & \text{pour } x \notin R_X \end{cases}$$

  - ses valeurs sont non négatives : $p(x) \geq 0 \ \forall x \in \mathbb{R}$
  - la somme des probabilités vaut 1 : $\sum_{x \in \mathbb{R}} p(x) = 1$
  - la probabilité d'un événement $B \subset R_X$ est : $P(B) = \sum_{x \in B} p(x)$

Exercice

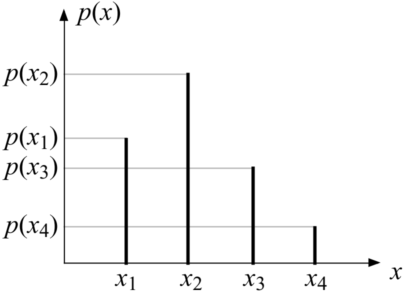
- On s'intéresse au nombre X d'insectes par plant de tabac. On donne les valeurs suivantes :

x	0	1	2	3	$\cdots$
$p(x)$	0.01	0.02	0.05	0.03	$\cdots$

- On demande de calculer :
  - ☐ $P(X < 3)$
  - ☐ $P(X \geq 2)$

Fonction de probabilité

- La fonction de probabilité $p(x)$ peut être définie :
  - par une table des probabilités $p(x) = P(X = x)$
  - par une formule mathématique (voir plus loin)
  - par un diagramme en bâtonnets



Fonction de répartition

- Fonction donnant $\forall x \in \mathbb{R}$ la probabilité $P(X \leq x)$ correspondante. On note en abrégé

$$F(x) = P(X \leq x) \qquad \forall x \in \mathbb{R}$$

  - ☐ elle prend ses valeurs sur $[0, 1]$: $0 \leq F(x) \leq 1 \ \forall x \in \mathbb{R}$
  - ☐ elle est monotone croissante : si $x_1 < x_2$, alors $F(x_1) \leq F(x_2)$
  - ☐ $F(\infty) = \lim_{x \to \infty} F(x) = 1$
  - ☐ $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$
  - ☐ pour $x_1 < x_2$, on a $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

Son allure générale est celle d'une fonction "en escalier"

Exemple

- On s'intéresse au nombre X d'insectes par plant de tabac.
  On donne les valeurs suivantes :

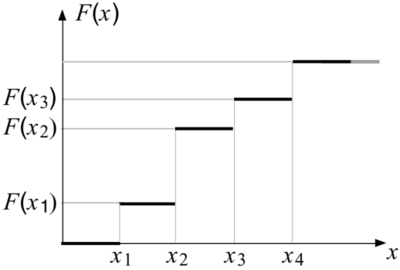
x	0	1	2	3	$\cdots$
$p(x)$	0.01	0.02	0.05	0.03	$\cdots$
$F(x)$					$\cdots$

- ☐ remplissez le tableau avec les valeurs pour $F(x)$
- ☐ sur base de ces valeurs, calculez :

$$P(0 < X \leq 2) \quad ; \quad P(X = 1)$$

Fonction de répartition

- La fonction de répartition $F(x)$ peut être définie :
  - par une table des probabilités $F(x) = P(X \leq x)$
  - par une formule mathématique (voir plus loin)
  - par un graphe "en escalier"



Loi de Bernouilli $X \sim Be(p)$

- De nombreuses expériences aléatoires peuvent être vues comme une répétition d'une expérience aléatoire binaire :

  Soient les événements A et A^* tels que $\Omega = \{A, A^*\}$

$$\left.\begin{array}{lll} - \text{ si } A & \text{est réalisé, on pose} & X = 1 \\ - \text{ si } A^* & \text{est réalisé, on pose} & X = 0 \end{array}\right\} \quad \Longrightarrow \quad R_X = \{0, 1\}$$

  Si $P(A) = p$, on a donc

$$p(x) = \begin{cases} p & \text{si } x = 1 \qquad (A \text{ réalisé}) \\ 1 - p & \text{si } x = 0 \qquad (A^* \text{ réalisé}) \\ 0 & \text{autrement} \end{cases}$$

Exemples

- Il est facile de trouver des exemples d'expériences dont le résultat est binaire : soit A est réalisé, soit A^* est réalisé.
  - Jeu de pile ou face : $A \equiv$"Pile" et $A^* \equiv$"Face"
  - Passage d'un examen : $A \equiv$"Réussi" et $A^* \equiv$"Raté"
  - Présence d'une maladie : $A \equiv$"Présente" et $A^* \equiv$"Absente"

- La loi de Bernouilli est importante car elle est une "brique élémentaire" pour la construction de variables aléatoires plus complexes.

Loi binomiale $X \sim Bi(n, p)$

- soit une expérience aléatoire telle que A et A^* sont possibles, avec $P(A) = p$ et $P(A^*) = 1 - p$.
- on répète n fois l'expérience de façon indépendante
- la probabilité de succès et d'échec reste identique
- la variable aléatoire X désigne le nombre de succès

La probabilité d'obtenir x succès est alors donnée par :

$$p(x) = \begin{cases} C_n^x p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, \ldots, n \\ 0 & \text{autrement} \end{cases}$$

avec $C_n^x = \dfrac{n!}{x!(n-x)!}$

Loi binomiale – propriétés

- pour $n = 1$, $X \sim Bi(n, p) \sim Be(p)$
- si $X_1 \sim Be(p), \ldots, X_n \sim Be(p)$ et $X_1 \perp \ldots \perp X_n$, alors

$$X_1 + \cdots + X_n \sim Bi(n, p)$$

- si $X \sim Bi(n, p)$, alors $Y = n - X \sim Bi(n, 1 - p)$
- formule de récurrence :

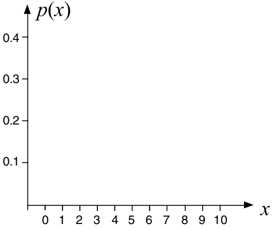
$$\begin{cases} p(0) = (1-p)^n \\ p(x) = p(x-1)\left(\dfrac{p}{1-p}\right)\left(\dfrac{n-x+1}{x}\right) \end{cases}$$

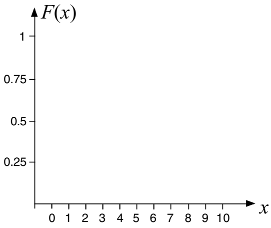
- utilisation de tables pour $p \leq 0.5$

Exercice

La fréquence théorique des individus porteurs d'un gène donné dans une population vaut 0.2. Si l'on prend 10 individus au hasard, calculez les valeurs suivantes :

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$											
$F(x)$											





Loi géométrique $X \sim Ge(p)$

- soit une expérience aléatoire telle que A et A^* sont possibles, avec $P(A) = p$ et $P(A^*) = 1 - p$.
- on répète l'expérience de façon indépendante
- la probabilité de succès et d'échec reste identique
- la variable X est le nombre de répétitions jusqu'au 1er succès

Les fonctions $p(x)$ et $F(x)$ sont données par :

$$p(x) = \begin{cases} p(1-p)^{x-1} & \text{si } x = 1, 2, \ldots \\ 0 & \text{autrement} \end{cases}$$

$$F(x) = 1 - (1-p)^x \quad x = 1, 2, \ldots$$

Loi géométrique – propriétés

- La loi géométrique prend une infinité dénombrable de valeurs
- Formule de récurrence :

$$\begin{cases} p(1) = p \\ \\ p(x+1) = p(x)(1-p) \end{cases}$$

- Le processus est sans mémoire : $P(X > k + x | X > k) = P(X > x)$

Exercice

- La probabilité qu'a un chasseur d'abattre un lapin qui court est de 0.6 à chaque coup tiré. Si le lapin est abattu, le chasseur ne tire plus. Quelle est la probabilité que le chasseur abatte le lapin :

  - ☐ en un coup ?
  - ☐ en deux coups ?
  - ☐ en moins de 4 coups ?
  - ☐ en deux coups sachant qu'il a déjà tiré 3 coups ?

Loi de Pascal $X \sim Pa(k, p)$

- soit une expérience aléatoire telle que A et A^* sont possibles, avec $P(A) = p$ et $P(A^*) = 1 - p$.
- on répète l'expérience de façon indépendante
- la probabilité de succès et d'échec reste identique
- la variable X est le nombre de répétitions jusqu'au $k$ème succès

La fonction $p(x)$ est donnée par :

$$p(x) = \begin{cases} C_{x-1}^{k-1} p^k (1-p)^{x-k} & \text{si } x = k, k+1, \ldots \\ 0 & \text{autrement} \end{cases}$$

avec $C_{x-1}^{k-1} = \dfrac{(x-1)!}{(k-1)!(x-k)!}$

Loi de Pascal – propriétés

- pour $k = 1$, $X \sim Pa(k, p) \sim Ge(p)$
- si $X_1 \sim Ge(p), \ldots, X_k \sim Ge(p)$ et $X_1 \perp \cdots \perp X_k$, alors

$$X_1 + \cdots + X_k \sim Pa(k, p)$$

- formule de récurrence :

$$\begin{cases} p(k) = p^k \\ p(x) = (1-p)\left(\dfrac{x-1}{x-k}\right) p(x-1) \quad \text{pour } x > k \end{cases}$$

Exercice

Une société prospecte pour l'exploitation d'un minerai en Afrique du Sud en sélectionnant des sites au hasard. Un site est considéré comme exploitable si la teneur en minerai dans au moins 8 des 10 échantillons prélevés sur ce site excède 30 % du poids brut. La probabilité d'excéder cette teneur pour un échantillon pris au hasard est de 70 %. On considère que les échantillons sont indépendants. La société décide de prospecter de manière à sélectionner 3 sites. Quelle est la probabilité qu'elle doivent examiner $3, 4, \ldots$ sites ?

Loi de Poisson $X \sim Po(\mu)$

- soit X le nombre de réalisations d'un événement A sur un support (intervalle de temps, surface, volume,...) de taille S.
- la probabilité de réalisation de A est très faible
- le nombre de réalisations possibles pour A est élevé

La probabilité d'observer x occurrences est donnée par :

$$p(x) = \frac{e^{-\mu}\mu^x}{x!}$$

Le paramètre μ (adimensionnel) est le produit d'une intensité λ par la taille S du support :

$$\mu = \lambda S$$

Exemples

- nombre de particules radioactives émises en un temps donné
- nombre de personnes arrivant à un guichet sur 10 minutes
- nombre d'erreurs typographiques sur une page d'un livre
- nombre de militaires tués par ruade de cheval sur une année (Von Bortkiewicz, 1879)
- nombre de V1 tombés sur un quartier de Londres

Exemple historique

Un quartier de Londres a subi un bombardement de V1 durant la 2ème guerre mondiale (537 impacts). On a divisé la zone bombardée en 576 carreaux de 500 x 500 m. On compte le nombre de carreaux ayant subi $x = 0, 1, 2, \ldots$ impacts

$$X \sim Po(\mu) \qquad \text{avec } \mu = 0.9288$$

x	0	1	2	3	4	5	$\cdots$
Nombre	229	211	93	35	7	1	$\cdots$
Fréquence	0.398	0.366	0.162	0.061	0.012	0.002	$\cdots$
$p(x)$	0.395	0.367	0.170	0.053	0.012	0.002	$\cdots$

La modélisation par une loi de Poisson est très satisfaisante.

Loi de Poisson – propriétés

- si $X_1 \sim Po(\mu_1), \ldots, X_k \sim Po(\mu_k)$ et $X_1 \perp \cdots \perp X_k$, alors

$$X_1 + \cdots + X_k \sim Po(\mu_1 + \cdots + \mu_k)$$

- Formule de récurrence :

$$\begin{cases} p(0) = e^{-\mu} \\ p(x) = p(x-1)\left(\frac{\mu}{x}\right) \end{cases}$$

- Utilisation de tables
- Approximation d'une loi $Bi(n, p)$ par une loi $Po(np)$:

$$\begin{cases} X \sim Bi(n,p) \\ n \to \infty, p \to 0 \\ np \text{ est une valeur finie} \end{cases} \implies \lim_{n \to \infty} p(x) = \frac{e^{-\mu}\mu^x}{x!}$$

Exercice

On suppose que le nombre de colonies d'une bactérie pathogène dans 10 ml de solution suit une loi de Poisson de paramètre $\mu = 20$. Quelle est la probabilité qu'il n'y ait pas de colonie dans :

- ☐ 1ml de solution ?
- ☐ 3 ml de solution ?

Exercice

Une banque possède 6000 clients. La probabilité que l'un d'entre eux se présente à un guichet entre 12h00 et 12h10 est de 0.0003. Quelle est la probabilité qu'il y ait entre 12h00 et 12h10 :

- ☐ aucun client au guichet ?
- ☐ un seul client au guichet ?
- ☐ deux clients au guichet ?

Loi hypergométrique $X \sim Hy(N, n, k)$

- soit un tirage sans remise de n objets parmi N objets
- parmi les N objets, k réalisent l'événement A
- la variable aléatoire X est le nombre d'objets tirés réalisant A

La probabilité d'obtenir x objets réalisant A est :

$$p(x) = \begin{cases} \dfrac{C_k^x C_{N-k}^{n-x}}{C_N^n} & \text{si } x = \max(0, n+k-N), \ldots, \min(n,k) \\ 0 & \text{autrement} \end{cases}$$

Quand $N >>$, on approxime $Hy(N, n, k)$ par $Bi(n, k/N)$.

Exercice

Dans un troupeau de 100 vaches, 12 sont malades. On sélectionne 5 vaches au hasard. Quelles sont les probabilités qu'il y ait x vaches malades, avec $x = 0, 1, \cdots, 5$?

Variable aléatoire continue

- Le nombre de valeurs possibles pour X est une infinité non dénombrable.

Exemples :

- ☐ proportion d'individus atteint d'une maladie
  $R_X =$
- ☐ durée de vie d'une ampoule électrique
  $R_X =$
- ☐ erreur de mesure d'une balance électronique
  $R_X =$

Variable aléatoire continue

- On peut caractériser une variable aléatoire continue par :
  - ☐ sa fonction de répartition $F(x) = P(X \leq x)$

    – mêmes propriétés que pour une variable discrète
    – fonction continue et non plus fonction en escalier

  - ☐ sa fonction de **densité** de probabilité $f(x)$

$$f(x) = \frac{dF(x)}{dx} \qquad F(x) = \int_{-\infty}^{x} f(t)dt$$

- La fonction $f(x)$ est exprimée à l'aide d'une formule mathématique. La fonction $F(x)$ n'a pas toujours d'expression analytique.

Note

- Pour une variable aléatoire$X$ continue, on a toujours

$$P(X = x) = 0 \quad \Longleftrightarrow \quad P(X < x) = P(X \leq x)$$

Exemple : Soit l'expérience aléatoire qui consiste à :

- choisir aléatoirement et indépendamment n chiffres de 0 à 9
- s'assurer que chaque chiffre est équiprobable.
- associer le $i$ème chiffre à la $i$ème décimale d'un nombre

si on tire 9 puis 7, on aura 0.97
si on tire 0 puis 2, on aura 0.02

On demande de calculer la probabilité qu'un nombre x donné parmi l'ensemble des nombres possibles compris entre 0 et 1 (1 étant exclu) apparaisse quand n vaut $1, 2, 3, \ldots$

Fonction de densité de probabilité

- Fonction donnant pour tout $x \in \mathbb{R}$ la dérivée de la fonction $F(x)$. On aura donc

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \quad \Longleftrightarrow \quad F(x) = \int_{-\infty}^{x} f(t)dt$$

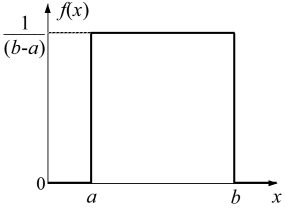
- Elle possède les propriétés suivantes :
  - ☐ ses valeurs sont non négatives : $f(x) \geq 0 \ \forall x \in \mathbb{R}$
  - ☐ son intégrale sur $\mathbb{R}$ vaut 1 : $\displaystyle\int_{-\infty}^{+\infty} f(t)dt = 1$
  - ☐ la probabilité d'un événement $B \subset R_X$ est : $P(B) = \displaystyle\int_{x \in B} f(x)dx$

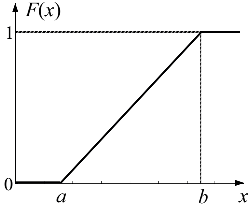
Loi uniforme $X \sim Un(a, b)$

- soit X un nombre réel sur l'intervalle $R_X = [a, b]$
- chaque valeur x a la même chance d'apparaître

Les fonctions $f(x)$ et $F(x)$ sont données par :

$$f(x) = \frac{1}{b-a} \qquad F(x) = \frac{x-a}{b-a} \quad \forall x \in [a, b]$$





Exemples

- Dans un casino, on fait tourner une roulette verticale munie d'un repère, qui freine lentement sous l'effet du frottement. L'angle X que forme le repère avec la verticale est distribué uniformément : $X \sim Un(0, 360)$

- Dans une station de métro, les rames arrivent à une cadence de 1 toutes les 5 minutes. Un passager se présente à un instant qui n'est lié d'aucune façon à l'horaire des rames. Le temps d'attente X est distribué uniformément : $X \sim Un(0, 5)$

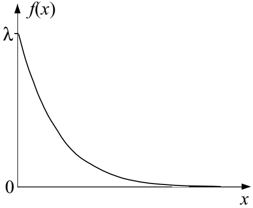
- On mesure la longueur en cm de feuilles de papier à l'aide d'une latte graduée en cm. L'erreur de mesure X en cm est distribuée uniformément : $X \sim Un(-0.5, 0.5)$

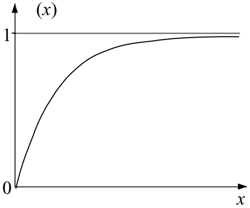
Loi exponentielle $X \sim Exp(\lambda)$

- soit X un nombre réel sur l'intervalle $R_X = [0, \infty)$
- le processus est sans mémoire : $P(X > k + x | X > k) = P(X > x)$

Les fonctions $f(x)$ et $F(x)$ sont données par :

$$f(x) = \lambda e^{-\lambda x} \qquad F(x) = 1 - e^{-\lambda x} \quad \forall x \in [0, \infty)$$





Loi exponentielle – propriétés

- Soit un processus se déroulant dans le temps.
  Soit $X \sim Exp(\lambda)$ le temps séparant deux occurrences.
  Si Y est le nombre d'occurrences sur un intervalle T, alors

$$Y \sim Po(\lambda T)$$

- Le processus est sans mémoire : savoir ce qui s'est déjà passé n'influencera pas le futur

$$P(X > k + x | X > k) = P(X > x)$$

Exemples

- On s'intéresse au temps X de fonctionnement d'un appareil entre deux pannes. Si le nombre Y de pannes sur une période T suit une distribution $Y \sim Po(\lambda T)$, alors $X \sim Exp(\lambda)$.
- On s'intéresse au temps X séparant l'arrivée de deux clients à un guichet. Si le nombre Y de clients arrivant au guichet sur une période T suit une distribution $Y \sim Po(\lambda T)$, alors $X \sim Exp(\lambda)$.
- On s'intéresse au temps X séparant deux catastrophes aériennes. Si le nombre Y de ces catastrophes sur une période T suit une distribution $Y \sim Po(\lambda T)$, alors $X \sim Exp(\lambda)$.

Exercice

Une banque possède 6000 clients. La probabilité qu'un client bien particulier se présente au guichet entre 12h00 et 12h10 est de 0.0003. Cette probabilité est identique d'un client à l'autre.

Un client se présente au guichet à 12h07. Quelle est la probabilité que le client suivant arrive avant 12h10 ?

Loi normale $X \sim N(\mu, \sigma^2)$

- soit X un nombre réel sur l'intervalle $R_X = (-\infty, \infty)$
- le processus est la somme d'effets issus d'un grand nombre de causes agissant indépendamment

☐ La fonction $f(x)$ est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x \in \mathbb{R}$$

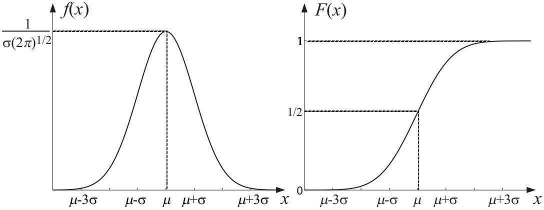
avec $-\infty < \mu < \infty$ et $\sigma^2 > 0$.

☐ La fonction $F(x)$ n'a pas d'expression analytique.

☐ Les valeurs de probabilités s'obtiennent par intégration numérique :

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

$$P(a \leq X \leq b) = \int_{a}^{b} f(t)dt$$



Loi normale – propriétés

- Distribution symétrique autour de μ
- Loi normale réduite (utilisation de tables) :

$$X \sim N(\mu, \sigma^2) \quad \Longrightarrow \quad \frac{X - \mu}{\sigma} \sim N(0, 1)$$

- La somme de lois normales (indépendantes ici) est normale

$$\begin{cases} X_i \sim N(\mu_i, \sigma_i^2) \\ X_1 \perp \cdots \perp X_k \end{cases} \Longrightarrow \quad \sum_{i=1}^{k} X_i \sim N(\sum_{i=1}^{k} \mu_i, \sum_{i=1}^{k} \sigma_i^2)$$

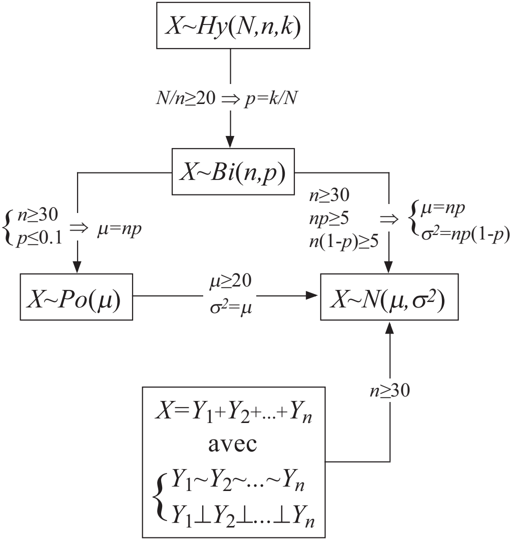
Exemples

- Production laitière d'une vache d'une race donnée
- Taille des individus dans une population donnée
- Machine de Galton
  www.ms.uky.edu/∼mai/java/stat/GaltonMachine.html
- Distribution d'une grandeur calculée comme une moyenne (théorème central limite)

Exercice

La production d'une vache laitière d'une race donnée suit une loi normale de paramètres μ=20 litres/jour et σ^2=9 litres^2/jour^2.
On demande de calculer :

- ☐ la probabilité qu'une vache produise moins de 12 litres/jour
- ☐ la probabilité qu'une vache produise plus de 23 litres/jour


$X \sim Hy(N,n,k)$
$N/n \geq 20 \Rightarrow p=k/N$
$X \sim Bi(n,p)$
$\begin{cases} n \geq 30 \\ p \leq 0.1 \end{cases} \Rightarrow \mu = np$
$n \geq 30$
$np \geq 5$
$n(1-p) \geq 5$
$\Rightarrow \begin{cases} \mu = np \\ \sigma^2 = np(1-p) \end{cases}$
$X \sim Po(\mu)$
$\mu \geq 20$
$\sigma^2 = \mu$
$X \sim N(\mu, \sigma^2)$
$n \geq 30$
$X = Y_1 + Y_2 + \ldots + Y_n$
avec
$\begin{cases} Y_1 \sim Y_2 \sim \ldots \sim Y_n \\ Y_1 \perp Y_2 \perp \ldots \perp Y_n \end{cases}$

Remarque

- Il existe un grand nombre d'autres lois de probabilité utiles :
  - Loi log-normale
  - Loi de Cauchy
  - Loi de Raleigh
  - Loi d'Erlang
  - Loi du χ^2
  - Loi de Student
  - Loi de Fisher-Snedecor

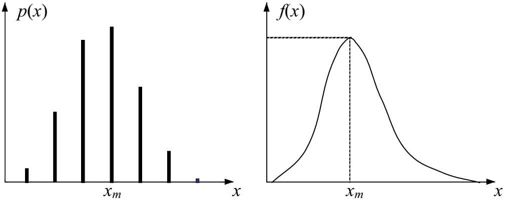
  $\vdots$

4. Grandeurs caractéristiques

Mode x_m

C'est la valeur la plus probable ou vraisemblable que peut prendre la variable aléatoire X.

$$p(x_m) > p(x) \quad \text{ou} \quad f(x_m) > f(x) \qquad \forall x \in \mathbb{R}, x \neq x_m$$

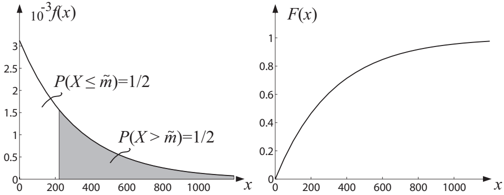


- le mode est le maximum de $p(x)$ ou de $f(x)$ sur son domaine.
- le mode n'est pas toujours défini (ex : lorsque $X \sim Un(a,b)$)

Médiane $\tilde{m}$

C'est la valeur coupant la distribution de X en deux intervalles de probabilités identiques.

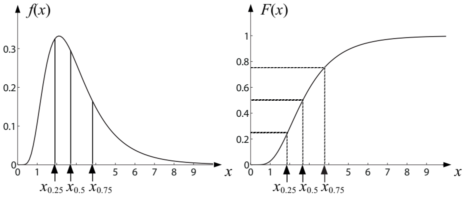
$$P(X \leq \tilde{m}) = F(\tilde{m}) = \frac{1}{2}$$



- la médiane n'est définie que si $\tilde{m} \in R_X$.
- pour toute fonction $f(x)$ symétrique autour de c, $\tilde{m} = c$.
- si $F(\tilde{m}) = \frac{1}{2}$, alors $\tilde{m} = F^{-1}(\frac{1}{2})$.

Quantile x_p

C'est la valeur x_p telle que $P(X \leq x_p) = F(x_p) = p$.



- la fonction $x_p = F^{-1}(p)$ est la fonction réciproque de $p = F(x_p)$
- il est d'usage de donner un nom particulier à certains quantiles
- les quantiles interviennent dans la notion d'intervalle de confiance

Quantile x_p

Quelques quantiles couramment utilisés sont :

- la médiane, qui est le 0.5 quantile : $\tilde{m} = x_{0.5}$
- les quartiles $x_{0.25}$, $x_{0.50}$ et $x_{0.75}$

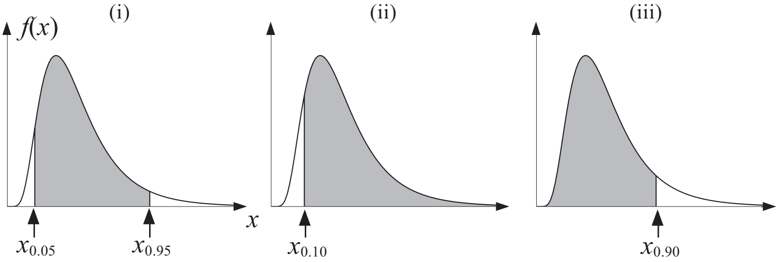
  $x_{0.25}$: premier quartile
  $x_{0.50}$: deuxième quartile (ou médiane)
  $x_{0.75}$: troisième quartile

- les déciles $x_{0.1}, x_{0.2}, \ldots, x_{0.9}$
- les centiles $x_{0.01}, x_{0.02}, \ldots, x_{0.99}$

Intervalle de confiance au niveau p

C'est un intervalle $[a, b]$ de valeurs tel que $P(a \leq X \leq b) = p$.
Les valeurs a et b sont finies ou infinies.

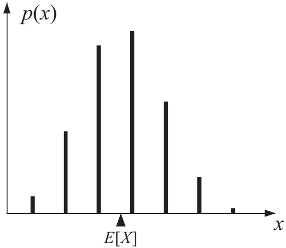
Exemple : intervalles de confiance au niveau $p = 0.9$

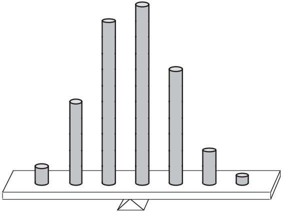


Espérance μ ou $E[X]$

C'est le centre de gravité de la fonction $p(x)$ ou $f(x)$.

$$\mu = E[X] = \begin{cases} \displaystyle\sum_{i=1}^{n} x_i p(x_i) & \text{cas discret} \\ \\ \displaystyle\int_{-\infty}^{+\infty} x f(x) dx & \text{cas continu} \end{cases}$$





Exercice

Un ami vous propose de jouer au dé de la façon suivante :

☐ si le résultat est 1 ou 2,	vous gagnez 100 Euros
☐ si le résultat est 3,	vous gagnez 200 Euros
☐ si le résultat est 4,	vous perdez 50 Euros
☐ si le résultat est 5,	vous perdez 100 Euros
☐ si le résultat est 6,	vous perdez 300 Euros

Avez-vous intérêt à parier avec votre ami ?

Remarques

- Espérance = valeur moyenne attendue (moyenne mathématique)
- Dans le cas discret : $E[X]$ est la moyenne pondérée des x_i
- L'espérance n'est pas toujours une valeur $\in R_X$
- L'espérance n'existe qu'à condition que :

$$\sum_{i=1}^{\infty} |x_i| p(x_i) < +\infty \qquad \int_{-\infty}^{+\infty} |x| f(x) dx < +\infty$$

(convergence absolue de la somme ou de l'intégrale)

- Ne pas confondre espérance, mode et médiane

Exercice

Sur base de sa longue expérience, un marchand de chaussures a pu établir la fréquence théorique des pointures x des paires de chaussures vendues à sa clientèle masculine adulte :

x	40	41	42	43	44	45	46
$p(x)$	0.04	0.16	0.28	0.24	0.16	0.08	0.04

Quelle est :

- ☐ la pointure la plus vendue ?
- ☐ la pointure moyenne des clients ?

Exercice

La durée de vie X d'une ampoule électrique suit une loi exponentielle de paramètre $\lambda = 1/200$ jours^{-1}. Quelle est la durée de vie moyenne d'une ampoule ?

Espérance des principales lois de probabilité

- Loi discrètes :
  - Bernouilli : $X \sim Be(p) \implies E[X] = p$
  - Binomiale : $X \sim Bi(n,p) \implies E[X] = np$
  - Géométrique : $X \sim Ge(p) \implies E[X] = 1/p$
  - Pascal : $X \sim Pa(k,p) \implies E[X] = k/p$
  - Poisson : $X \sim Po(\mu) \implies E[X] = \mu$

- Lois continues :
  - uniforme : $X \sim Un(a,b) \implies E[X] = (a+b)/2$
  - exponentielle : $X \sim Exp(\lambda) \implies E[X] = 1/\lambda$
  - normale : $X \sim N(\mu,\sigma^2) \implies E[X] = \mu$

Variance σ^2 ou $Var[X]$

C'est une mesure de la dispersion d'une variable aléatoire.

$$\sigma^2 = Var[X] = \begin{cases} \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i) & \text{cas discret} \\ \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx & \text{cas continu} \end{cases}$$

Des formules équivalentes sont :

$$Var[X] = \left(\sum_{i=1}^{n} x_i^2 p(x_i) \right) - \mu^2 \qquad Var[X] = \int_{-\infty}^{+\infty} x^2 f(x) dx - \mu^2$$

- ses unités sont le carré des unités de la variable
- la valeur $\sigma = \sqrt{\sigma^2}$ est appelée écart-type ou déviation standard

Remarques

- Variance = moment d'inertie de la distribution
- Cas discret : $Var[X]$ est la moyenne pondérée des $(x_i - \mu)^2$
- La variance n'existe que s'il y a convergence absolue de la somme ou de l'intégrale :

$$\sum_{i=1}^{\infty}(x_i - \mu)^2 p(x_i) < +\infty \qquad \int_{-\infty}^{+\infty}(x_-\mu)^2 f(x)dx < +\infty$$

Variance des principales lois de probabilité

- Loi discrètes :
  - Bernouilli : $X \sim Be(p) \implies Var[X] = p(1-p)$
  - Binomiale : $X \sim Bi(n,p) \implies Var[X] = np(1-p)$
  - Géométrique : $X \sim Ge(p) \implies Var[X] = (1-p)/p^2$
  - Pascal : $X \sim Pa(k,p) \implies Var[X] = k(1-p)/p^2$
  - Poisson : $X \sim Po(\mu) \implies Var[X] = \mu$
- Lois continues :
  - uniforme : $X \sim Un(a,b) \implies Var[X] = (b-a)^2/12$
  - exponentielle : $X \sim Exp(\lambda) \implies Var[X] = 1/\lambda^2$
  - normale : $X \sim N(\mu,\sigma^2) \implies Var[X] = \sigma^2$

5. Fonction d'une variable

Introduction

Soient X une variable aléatoire et $Y = h(X)$ une fonction de X.
Comment obtenir

- ☐ la loi de probabilité de Y ?
- ☐ l'espérance et la variance de Y ?

Réponse : cela dépend de la forme de la fonction $h(.)$;
on distingue ainsi les cas pour lesquels :

$h(.)$ est quelconque
$h(.)$ est strictement monotone
$h(.)$ est linéaire
$h(.)$ est linéarisable

Exemples

- Soit X l'erreur de mesure angulaire (en degrés) sur la hauteur d'un objet placé à une distance connue, avec $X \sim Un(-\frac{1}{2}, \frac{1}{2})$.

  Si Y est l'erreur sur la hauteur de l'objet, quelle sera sa distribution ?

- Soit X le diamètre (en cm) d'une bille fabriquée par une machine. On sait que $X \sim Un(4, 4.1)$.

  Si Y est son volume, quelle sera sa distribution ?

Fonction $h(.)$ quelconque (cas général)

- On identifie des événements A et B équivalents ($A \equiv B$) sur les variables $Y = h(X)$ et X.

  - Cas discret :

$$\begin{cases} A \equiv (Y = y) \\ \\ B \equiv \{x : h(x) = y\} \end{cases} \Longrightarrow \quad p_Y(y) = \sum_{x \in B} p_X(x)$$

  - Cas continu :

$$\begin{cases} A \equiv (Y \leq y) \\ \\ B \equiv \{x : h(x) \leq y\} \end{cases} \Longrightarrow \quad F_Y(y) = \int_{x \in B} f_X(t)dt$$

Exemple

Soit X une variable aléatoire, avec $X \sim Un(-1, 1)$.
Soit la nouvelle variable $Y = X^2$.

Comment obtenir $f_Y(y)$ à partir de $f_X(x)$?

Fonction $h(.)$ strictement monotone

- Pour $h(.) \nearrow$ ou $\searrow$, on a toujours

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right|$$

La fonction $h^{-1}(.)$ est la réciproque de $h(.)$

Exemple

Soit X le diamètre (en cm) de graines sphériques, où $X \sim Un(1, 3)$.
Soit Y le volume de ces graines.

Comment obtenir $f_Y(y)$ à partir de $f_X(x)$?

Fonctions $h(.)$ linéaires

- Soit une fonction linéaire

$$Y = h(X) = aX + b$$

C'est un cas particulier des fonctions strictement monotones :

$$x = h^{-1}(y) = \frac{y-b}{a} \quad \Longrightarrow \quad \left|\frac{dh^{-1}(y)}{dy}\right| = \left|\frac{1}{a}\right|$$

On obtient donc

$$f_Y(y) = f_X(\frac{y-b}{a}) \left|\frac{1}{a}\right|$$

Espérance d'une variable $Y = h(X)$

Pour calculer $E[Y]$, on peut :

- ☐ déterminer $f_Y(y)$ ou $p_Y(y)$ à partir de $f_X(x)$ ou $p_X(x)$
- ☐ calculer $E[Y]$ sur base des formules classiques

En pratique, on montre qu'un calcul plus rapide est :

$$E[Y] = E[h(X)] = \begin{cases} \displaystyle\sum_{i=1}^{n} h(x_i)p(x_i) & \text{cas discret} \\ \displaystyle\int_{-\infty}^{+\infty} h(x)f(x)dx & \text{cas continu} \end{cases}$$

Exemple

Soit X le diamètre (en cm) de graines sphériques, où $X \sim Un(1,3)$.
Soit Y le volume de ces graines.

Comment obtenir $E[Y]$?

Espérance d'une fonction linéaire

On montre facilement que si $Y = aX + b$, alors

$$E[Y] = aE[X] + b$$

- L'espérance est donc un opérateur linéaire
- De façon générale, $E[h(X)] \neq h(E[X])$

Exemple

On envisage de creuser un puits pour atteindre une nappe phréatique, dont la profondeur est en moyenne égale à 30 mètres dans la région. Le coût C du pompage se composera d'un coût fixe de 0.30 euros par mètre cube pompé (amortissement de l'installation, taxes, etc.) et d'un coût variable qui dépendra de la profondeur D du pompage ; il est de $0.03D$ euros par mètre cube pompé.

A quel coût par mètre cube pompé doit-on s'attendre pour un tel puits ?

Remarque

La variance peut être exprimée sur base d'espérances de fonctions de la variable X :

$$Var[X] = \begin{cases} \displaystyle\sum_{i=1}^{n}(x_i - \mu)^2 p(x_i) \\ \displaystyle\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx \end{cases} \iff Var[X] = E[(X-\mu)^2]$$

$$Var[X] = \begin{cases} \left(\displaystyle\sum_{i=1}^{n} x_i^2 p(x_i)\right) - \mu^2 \\ \displaystyle\int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2 \end{cases} \iff Var[X] = E[X^2] - E^2[X]$$

Variance d'une variable $Y = h(X)$

Pour calculer $Var[Y]$, on a les formules :

$$Var[Y] = \begin{cases} E[(Y-\mu_Y)^2] \\ \\ E[Y^2] - E^2[Y] \end{cases}$$

Il suffit donc de calculer les espérances correspondantes, avec

$$\mu_Y = E[Y] = E[h(X)] \qquad E[Y^2] = E[h^2(X)]$$

où $h(X)$ et $h^2(X)$ sont des fonctions de X.

Variance d'une fonction linéaire

On montre facilement que si $Y = aX + b$, alors

$$Var[Y] = a^2 Var[X]$$

- La variance n'est pas un opérateur linéaire
- La constante b n'affecte pas la variance.

Linéarisation de fonctions non linéaires

Soit X une variable aléatoire.
Soit $Y = h(X)$ une fonction de cette variable.

Calculer $E[Y]$ et $Var[Y]$ est

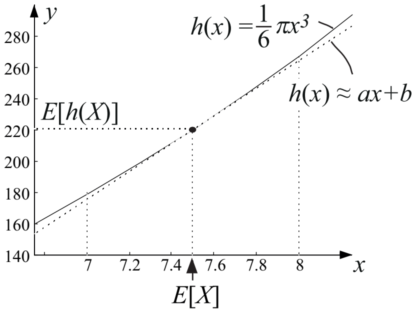
- ☐ laborieux si $h(.)$ est quelconque
- ☐ facile si $h(.)$ est linéaire

Ne pourrait-on pas approximer $h(.)$ par une fonction linéaire ?

Exemple

Soit $X \sim Un(7,8)$ le diamètre d'oranges calibrées.
On s'intéresse au volume $Y = \frac{1}{6}\pi X^3$ des oranges.

Quelles sont de bonnes approximations pour $E[Y]$ et $Var[Y]$?



6. Couples aléatoires

Couple aléatoire

Soient X et Y des variables aléatoires.
On dira que (X, Y) forme un couple aléatoire, caractérisé par :

- ☐ son domaine de variation conjoint
- ☐ sa fonction de répartition conjointe
- ☐ sa fonction de (densité de) probabilité conjointe

Le couple aléatoire est un cas particulier du vecteur aléatoire.
On considère ici $n = 2$ variables en même temps.

Exemples

- On s'intérèsse à la température minimale X et maximale Y d'un jour donné de l'année. Ces deux variables sont "liées" l'une à l'autre (quand X est grand, Y aura tendance à être grand et vice-versa).

- On s'intérèsse à la taille X et au poids Y d'un homme issu de la population belge. Ces deux variables sont "liées" l'une à l'autre (quand X est grand, Y aura tendance à être grand et vice-versa).

$\Longrightarrow$ Les deux variables aléatoires ne sont par liées par une relation fonctionnelle, mais connaître la valeur prise par l'une d'elles apporte de l'information sur les valeurs plausibles pour l'autre.

Domaine de variation R_{XY}

Le domaine R_{XY} est un sous-ensemble de $R_X \times R_Y$ constitué des couples de valeurs (x, y) observables conjointement.

Exemple :

Soit X et Y les pourcentages de sable et de limon dans un sol.
Que sont R_X, R_Y et R_{XY} ?

Fonction $F(x, y)$

- fonction donnant la probabilité de réalisation simultanée des événements $(X \leq x)$ et $(Y \leq y)$

$$F(x, y) = P(X \leq x \cap Y \leq y)$$

- $F(x, y)$ prend ses valeurs sur $[0, 1]$:

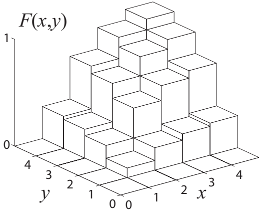
$$F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0$$

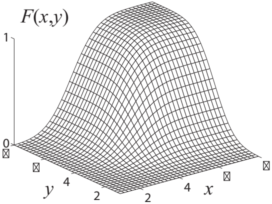
$$F(\infty, \infty) = 1$$

- $F(x, y)$ est monotone croissante :

$$F(x_i, y_i) \leq F(x_j, y_j) \qquad \forall x_i < x_j, \forall y_i < y_j$$

Exemple de graphes pour $F(x, y)$





- ☐ cas discret : $F(x, y)$ est une fonction "en marches d'escalier"
- ☐ cas continu : $F(x, y)$ est une fonction continue

Fonction $p(x, y)$:

- fonction donnant la probabilité de réalisation simultanée des événements $(X = x)$ et $(Y = y)$ (cas de variables X et Y discrètes)

$$p(x,y) = \begin{cases} P(X = x \cap Y = y) & \text{si } (x,y) \in R_{XY} \\ 0 & \text{si } (x,y) \notin R_{XY} \end{cases}$$

- ☐ la fonction $p(x, y)$ ne prend que des valeurs positives ou nulles :

$$p(x,y) \geq 0 \qquad \forall (x,y) \in \mathbb{R}^2$$

- ☐ la somme des valeurs de la fonction $p(x, y)$ est égale à l'unité :

$$\sum_i \sum_j p(x_i, y_j) = 1 \qquad (x_i, y_j) \in R_{XY}$$

Fonction $f(x,y)$:

- dérivée par rapport à x et y de la fonction $F(x,y)$ (cas de variables X et Y continues)

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t,u)dudt$$

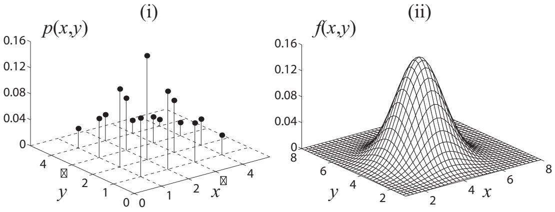
  - ☐ la fonction $f(x,y)$ ne prend que des valeurs positives ou nulles :

$$f(x,y) \geq 0 \qquad \forall (x,y) \in \mathbb{R}^2$$

  - ☐ la double intégrale de la fonction $f(x,y)$ sur $\mathbb{R}^2$ vaut 1 :

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dydx = 1$$

Exemple de graphes pour $p(x, y)$ *et* $f(x, y)$



- ☐ cas discret : $p(x, y)$ est une fonction “en bâtonnets”
- ☐ cas continu : $f(x, y)$ est une fonction continue

Probabilité d'un événement B

- Soit $B \subseteq R_{XY}$ un événement quelconque.
  Sa probabilité est donnée par

$$P(B) = \begin{cases} \displaystyle\sum_{(x,y)\in B} p(x,y) & \text{cas discret} \\ \\ \displaystyle\int_{(x,y)\in B} f(x,y)dydx & \text{cas continu} \end{cases}$$

  - ☐ cas discret : somme des probabilités $p(x,y)$ pour $(x,y) \in B$
  - ☐ cas continu : intégrale de la fonction $f(x,y)$ sur le domaine B

Exemple (cas discret)

On s'intéresse au nombre de chats X et de chiens Y que possède un ménage. On donne les probabilités $p(x, y)$ suivantes :

$x \backslash y$	0	1	2	3
0	0.6667	0.1333	0.0267	0.0067
1	0.0667	0.0133	0.0027	0.0007
2	0.0444	0.0089	0.0018	0.0004
3	0.0222	0.0044	0.0009	0.0002

- ☐ quelle est la fonction $F(x, y)$?
- ☐ quelle est la probabilité qu'il y ait plus de chats que de chiens ?
- ☐ quelle est la probabilité qu'il y ait plus d'un chat ?

Exemple (cas continu)

Soient X et Y les pourcentages de sable et de limon dans un type de sol donné. On suppose que $f(x, y) = k$ sur le domaine des valeurs possibles.

- calculez la valeur de k
- calculez la probabilité qu'il y ait plus de limon que de sable

Distributions marginales

- distribution ne dépendant plus que d'une des 2 variables
  - fonctions de répartition :

$$F_X(x) = P(X \leq x) = F(x, \infty)$$

$$F_Y(y) = P(Y \leq y) = F(\infty, y)$$

  - fonctions de probabilité :

$$p_X(x) = \sum_j p(x, y_j) \qquad p_Y(y) = \sum_i p(x_i, y)$$

  - fonctions de densité de probabilité :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \qquad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

Exemple (cas discret)

On s'intéresse au nombre de chats X et de chiens Y que possède un ménage. On donne les probabilités $p(x, y)$ suivantes :

$x \backslash y$	0	1	2	3
0	0.6667	0.1333	0.0267	0.0067
1	0.0667	0.0133	0.0027	0.0007
2	0.0444	0.0089	0.0018	0.0004
3	0.0222	0.0044	0.0009	0.0002

- ☐ quelles sont les fonctions $p_X(x)$ et $p_Y(y)$?
- ☐ quelles sont les fonctions $F_X(x)$ et $F_Y(y)$?

Exemple (cas continu)

Soient X et Y les pourcentages de sable et de limon dans un sol.
On suppose que $f(x, y) = 2$ sur R_{XY}.

- quelles sont les fonctions $f_X(x)$ et $f_Y(y)$?
- quelles sont les fonctions $F_X(x)$ et $F_Y(y)$?

Distributions conditionnelles

- on s'intéresse à la distribution d'une des variables connaissant les valeurs prises par l'autre variable.

(1) *Cas discret :*

☐ soient X et Y deux variables aléatoires
☐ soient $A \equiv (X = x)$ et $B \equiv (Y = y)$

Les fonctions de probabilité conditionnelles sont :

$$p_X(x|y) = P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{p(x,y)}{p_Y(y)} \quad \text{avec } p_Y(y) > 0$$

$$p_Y(y|x) = P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{p(x,y)}{p_X(x)} \quad \text{avec } p_X(x) > 0$$

Distributions conditionnelles

On définit aussi les fonctions de répartition conditionnelles :

$$F_X(x|y) = \sum_{x_i \leq x} p_X(x_i|y) \qquad F_Y(y|x) = \sum_{y_j \leq y} p_Y(y_j|x)$$

(2) *Cas continu :*

□ les fonctions de densité de probabilité conditionnelles sont :

$$f_X(x|y) = \frac{f(x,y)}{f_Y(y)} \quad \text{avec } f_Y(y) > 0$$

$$f_Y(y|x) = \frac{f(x,y)}{f_X(x)} \quad \text{avec } f_X(x) > 0$$

□ les fonctions de répartition conditionnelles sont :

$$F_X(x|y) = \int_{-\infty}^{x} f_X(t|y)dt \qquad F_Y(y|x) = \int_{-\infty}^{y} f_Y(u|x)du$$

Exemple (cas discret)

On s'intéresse au nombre de chats X et de chiens Y que possède un ménage. On donne les probabilités $p(x, y)$ suivantes :

$x \backslash y$	0	1	2	3
0	0.6667	0.1333	0.0267	0.0067
1	0.0667	0.0133	0.0027	0.0007
2	0.0444	0.0089	0.0018	0.0004
3	0.0222	0.0044	0.0009	0.0002

- ☐ quelles sont les valeurs de $p_X(x|y)$ pour $y = 0$?
- ☐ quelles sont les valeurs de $p_Y(y|x)$ pour $x = 1$?

Exemple (cas continu)

Soient X et Y les pourcentages de sable et de limon dans un sol. On suppose que $f(x, y) = 2$ sur R_{XY}.

- quelle est l'expression de la fonction $f_X(x|y)$?
- quelle est l'expression de la fonction $f_Y(y|x)$?

Indépendance

Soient X et Y deux variables aléatoires.
L'indépendance $X \perp Y$ est définie par les relations :

$$X \perp Y \iff \begin{cases} p_X(x|y) = p_X(x) & \forall x \in \mathbb{R} \\ p_Y(y|x) = p_Y(y) & \forall y \in \mathbb{R} \end{cases} \qquad \text{cas discret}$$

$$X \perp Y \iff \begin{cases} f_X(x|y) = f_X(x) & \forall x \in \mathbb{R} \\ f_Y(y|x) = f_Y(y) & \forall y \in \mathbb{R} \end{cases} \qquad \text{cas continu}$$

La connaissance de la valeur prise par l'une des variables n'affecte pas la distribution de l'autre variable.

De manière équivalente, on peut écrire

$$X \perp Y \Longleftrightarrow \begin{cases} f(x,y) = f_X(x) f_Y(y) \\ p(x,y) = p_X(x) p_Y(y) \end{cases} \qquad \forall (x,y) \in \mathbb{R}^2$$

Exercice : Pour les domaines de variation suivants, $X \perp Y$?

Covariance $Cov[X, Y]$

■ mesure de l'association linéaire entre X et Y

$$\sigma_{XY} = Cov[X, Y] = \begin{cases} \sum_i \sum_j (x_i - \mu_X)(y_j - \mu_Y) p(x_i, y_j) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y) f(x, y) dy dx \end{cases}$$

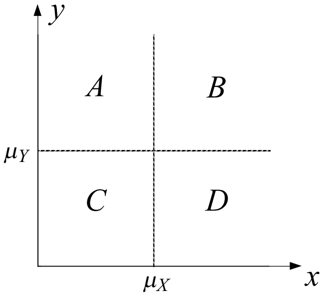
☐ des formules équivalentes sont :

$$\sigma_{XY} = \begin{cases} \left(\sum_i \sum_j x_i y_j p(x_i, y_j) \right) - \mu_X \mu_Y \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) dy dx - \mu_X \mu_Y \end{cases}$$

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]$$

☐ Ses unités sont le produit des unités des deux variables

Illustration



- ☐ lorsque $\sigma_{XY} > 0$, on parle d'association **linéaire** positive
- ☐ lorsque $\sigma_{XY} < 0$, on parle d'association **linéaire** négative
- ☐ lorsque $\sigma_{XY} = 0$, on parle d'absence d'association **linéaire**

Remarques

- $X \perp Y \Longrightarrow \sigma_{XY} = 0$ mais $\sigma_{XY} = 0 \not\Longrightarrow X \perp Y$
- On a toujours $-\sigma_X \sigma_Y \leq \sigma_{XY} \leq \sigma_X \sigma_Y$
- coefficient de corrélation : mesure standardisée de la covariance

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad -1 \leq \rho_{XY} \leq 1$$

$$X \perp Y \Longrightarrow \rho_{XY} = 0$$

Exemple (cas continu)

Soient X et Y les pourcentages de sable et de limon dans un sol.
On suppose que $f(x, y) = 2$ sur R_{XY}.

- quelle est la covariance entre X et Y ?
- quelle est la corrélation entre X et Y ?

Les définitions sont identiques à celles déjà vues, mais en utilisant les distributions conditionnelles.

- *Espérances conditionnelles :*

$$\mu_{Y|x} = E[Y|x] = \begin{cases} \sum_j y_j p_Y(y_j|x) \\ \int_{-\infty}^{+\infty} y f_Y(y|x) dy \end{cases}$$

$$\mu_{X|y} = E[X|y] = \begin{cases} \sum_i x_i p_X(x_i|y) \\ \int_{-\infty}^{+\infty} x f_X(x|y) dx \end{cases}$$

  - □ la fonction $E[Y|x]$ est aussi appelée régression de Y sur x
  - □ la fonction $E[X|y]$ est aussi appelée régression de X sur y.

Espérances et variances conditionnelles

- *variances conditionnelles :*

$$\sigma^2_{Y|x} = Var[Y|x] = \begin{cases} \sum_j (y_j - \mu_{Y|x})^2 p_Y(y_j|x) \\ \int_{-\infty}^{+\infty} (y - \mu_{Y|x})^2 f_Y(y|x) dy \end{cases}$$

$$\sigma^2_{X|y} = Var[X|y] = \begin{cases} \sum_i (x_i - \mu_{X|y})^2 p_X(x_i|y) \\ \int_{-\infty}^{+\infty} (x - \mu_{X|y})^2 f_X(x|y) dx \end{cases}$$

- Des formules équivalentes sont :

$$Var[Y|x] = E\left[(Y|x - \mu_{Y|x})^2\right] = E[Y^2|x] - \mu^2_{Y|x}$$

$$Var[X|y] = E\left[(X|y - \mu_{X|y})^2\right] = E[X^2|y] - \mu^2_{X|y}$$

Exemple (cas continu)

Soient X et Y les pourcentages de sable et de limon dans un sol.
On suppose que $f(x, y) = 2$ sur R_{XY}. Tracez les graphes de :

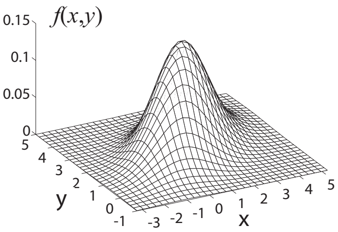
- la régression de X sur y
- la régression de Y sur x

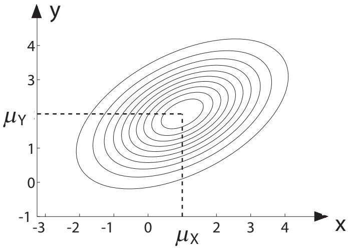
Couple aléatoire normal $(X, Y) \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$

Soient X, Y deux variables aléatoires continues
On parlera de couple (X, Y) normal lorsque

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

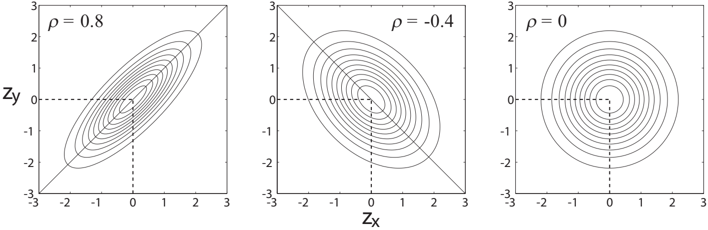
avec $(x, y) \in \mathbb{R}^2$, $\sigma_X > 0$, $\sigma_Y > 0$ et $-1 < \rho < 1$.





Influence de $\sigma_X^2, \sigma_Y^2, \rho$ sur la fonction $f(x,y)$

- ☐ les courbes de niveaux sont des ellipses centrées en (μ_X, μ_Y)
- ☐ la forme de l'ellipse dépend des paramètres $\sigma_X^2, \sigma_Y^2, \rho$



- ☐ la fonction de répartition $F(x,y)$ n'a pas d'expression analytique

Distributions marginales & conditionnelles

- Les distributions marginales sont des lois normales

$$(X,Y) \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho) \Longrightarrow \begin{cases} X \sim N(\mu_X, \sigma_X^2) \\ Y \sim N(\mu_Y, \sigma_Y^2) \end{cases}$$

Note : La relation réciproque n'est pas vraie !

- Les distributions conditionnelles sont des lois normales

$$(X,Y) \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho) \Longrightarrow \begin{cases} Y|x \sim N(\mu_{Y|x}, \sigma_{Y|x}^2) \\ X|y \sim N(\mu_{X|y}, \sigma_{X|y}^2) \end{cases}$$

- Les espérances conditionnelles sont données par

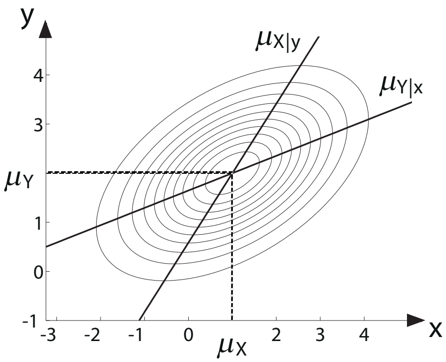
$$\begin{cases} E[Y|x] = \mu_{Y|x} = \mu_Y + \rho \dfrac{\sigma_Y}{\sigma_X}(x - \mu_X) \\ \\ E[X|y] = \mu_{X|y} = \mu_X + \rho \dfrac{\sigma_X}{\sigma_Y}(y - \mu_Y) \end{cases}$$

- Les variances conditionnelles sont données par

$$\begin{cases} Var[Y|x] = \sigma^2_{Y|x} = \sigma^2_Y(1 - \rho^2) \\ \\ Var[X|y] = \sigma^2_{X|y} = \sigma^2_X(1 - \rho^2) \end{cases}$$

Espérances & variances conditionnelles

- les régressions de X sur y et de Y sur x sont des droites
- lorsque $\rho = 0$, les régressions sont perpendiculaires



Indépendance

- pour un couple (X, Y) quelconque :

$$X \perp Y \Longrightarrow \rho = 0 \qquad \text{mais} \qquad \rho = 0 \not\Longrightarrow X \perp Y$$

- pour un couple (X, Y) normal :

$$X \perp Y \Longleftrightarrow \rho = 0$$

$\Longrightarrow$ Une corrélation nulle est synonyme d'indépendance

Exemple

On fait deux mesures de la pollution atmosphérique au même endroit.
La distribution conjointe forme un couple normal (X_1, X_2) avec :

- $\mu_1 = \mu_2 = 7$ mg/m^3
- $\sigma_1^2 = 3$ (mg/m^3)^2 et $\sigma_2^2 = 5$ (mg/m^3)^2
- $\rho = 0.9$

Soit un seuil d'alerte qui est de 10 mg/m^3.
On demande de calculer la probabilité que X_1 soit :

- au-delà du seuil d'alerte
- au-delà du seuil d'alerte sachant que $X_2 = 7$ mg/m^3
- au-delà du seuil d'alerte sachant que $X_2 = 12$ mg/m^3

7. Vecteurs aléatoires

Introduction

- le vecteur aléatoire est une généralisation du couple aléatoire.
    - Le vecteur des variables $X_1, X_2, \ldots, X_n$ est noté

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

    - Le vecteur des valeurs $x_1, x_2, \ldots, x_n$ est noté

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Fonction de (densité de) probabilité conjointe

- La distribution conjointe est donnée par $p(\mathbf{x})$ ou $f(\mathbf{x})$:

$$p(\mathbf{x}) = p(x_1, \ldots, x_n) \qquad f(\mathbf{x}) = f(x_1, \ldots, x_n)$$

- Les propriétés sont similaires à celles pour $p(x, y)$ et $f(x, y)$:

$$\begin{cases} p(\mathbf{x}) \geq 0 \ \forall \ \mathbf{x} \in \mathbb{R}^n \\ \\ \displaystyle\sum_{\mathbf{x}} p(\mathbf{x}) = \sum_{x_1} \cdots \sum_{x_n} p(x_1, \ldots, x_n) = 1 \end{cases}$$

$$\begin{cases} f(\mathbf{x}) \geq 0 \ \forall \ \mathbf{x} \in \mathbb{R}^n \\ \\ \displaystyle\int_{-\infty}^{+\infty} f(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_n) dx_n \cdots dx_1 = 1 \end{cases}$$

Fonction de répartition conjointe

- La fonction de répartition conjointe est notée

$$F(\mathbf{x}) = P(X_1 \leq x_1 \cap \cdots \cap X_n \leq x_n)$$

- cette fonction a des propriétés similaires à $F(x, y)$:

  □ si $x_i = +\infty \ \forall i = 1, \ldots, n$, alors $F(\mathbf{x}) = P(\Omega) = 1$
  □ si $x_i = -\infty \ \exists i = 1, \ldots, n$, alors $F(\mathbf{x}) = P(\emptyset) = 0$

- dans le cas d'un vecteur $\mathbf{X}$ de variables continues :

$$f(\mathbf{x}) = \frac{\partial^n F(\mathbf{x})}{\partial x_1 \cdots \partial x_n} \qquad F(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(\mathbf{t}) dt_n \cdots dt_1$$

Probabilité d'un événement

- La probabilité d'un événement B quelconque est donnée par

$$P(B) = \begin{cases} \displaystyle\sum_{\mathbf{x} \in B} p(\mathbf{x}) & \text{cas discret} \\ \\ \displaystyle\int_{\mathbf{x} \in B} f(\mathbf{x}) d\mathbf{x} & \text{cas continu} \end{cases}$$

La somme (l'intégrale) porte sur tous les vecteurs $\mathbf{x}$ qui réalisent B. Les sommes (les intégrales) sont n-variées.

Distributions marginales

- On peut définir une partition du vecteur $\mathbf{X}$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_{n_a} \\ \hline X_{n_a+1} \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_a \\ \mathbf{X}_b \end{pmatrix}$$

- Fonctions de (densité de) probabilité marginales

$$p(\mathbf{x}_a) = \sum_{\mathbf{x}_b} p(\mathbf{x}) = \sum_{x_{n_a+1}} \cdots \sum_{x_n} p(x_1, \ldots, x_n)$$

$$f(\mathbf{x}_a) = \int_{-\infty}^{+\infty} f(\mathbf{x}) d\mathbf{x}_b = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_n) dx_n \cdots dx_{n_a+1}$$

Distributions conditionnelles

- Fonctions de (densité de) probabilité conditionnelles

$$p(\mathbf{x}_a|\mathbf{x}_b) = \frac{p(\mathbf{x})}{p(\mathbf{x}_b)} \qquad f(\mathbf{x}_a|\mathbf{x}_b) = \frac{f(\mathbf{x})}{f(\mathbf{x}_b)}$$

- Théorème des probabilité totales

$$p(\mathbf{x}_a) = \sum_{\mathbf{x}_b} p(\mathbf{x}_a|\mathbf{x}_b)p(\mathbf{x}_b) \qquad f(\mathbf{x}_a) = \int_{-\infty}^{+\infty} f(\mathbf{x}_a|\mathbf{x}_b)f(\mathbf{x}_b)d\mathbf{x}_b$$

- Théorème des probabilité composées

$$f(x_1,\ldots,x_n) = f(x_1)f(x_2|x_1)f(x_3|(x_1,x_2))\cdots f(x_n|(x_1,\ldots,x_{n-1}))$$

$$p(x_1,\ldots,x_n) = p(x_1)p(x_2|x_1)p(x_3|(x_1,x_2))\cdots p(x_n|(x_1,\ldots,x_{n-1}))$$

- Théorème de Bayes

$$p(\mathbf{x}_b|\mathbf{x}_a) = \frac{p(\mathbf{x}_a|\mathbf{x}_b)p(\mathbf{x}_b)}{\sum_{\mathbf{x}_b} p(\mathbf{x}_a|\mathbf{x}_b)p(\mathbf{x}_b)} \qquad \text{cas discret}$$

$$f(\mathbf{x}_b|\mathbf{x}_a) = \frac{f(\mathbf{x}_a|\mathbf{x}_b)f(\mathbf{x}_b)}{\int_{-\infty}^{+\infty} f(\mathbf{x}_a|\mathbf{x}_b)f(\mathbf{x}_b)d\mathbf{x}_b} \qquad \text{cas continu}$$

Indépendance mutuelle

- Pour des variables $X_1, \ldots, X_n$ mutuellement indépendantes :

$$X_1 \perp \cdots \perp X_n \iff F(x_1, x_2, \ldots, x_n) = F(x_1)F(x_2)\cdots F(x_n)$$

$$X_1 \perp \cdots \perp X_n \iff \begin{cases} p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\cdots p(x_n) \\ \\ f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n) \end{cases}$$

- Il y a d'autres formes d'indépendance (partielle, conditionnelle,...)

Espérances, variances et covariances

- Vecteur espérance

$$\boldsymbol{\mu} = E[\mathbf{X}] = \begin{pmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_n} \end{pmatrix}$$

- Matrice de covariance

$$\Sigma = Cov[\mathbf{X}] = \begin{pmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_n} \\ \sigma_{X_2X_1} & \sigma^2_{X_2} & \cdots & \sigma_{X_2X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_nX_1} & \sigma_{X_nX_2} & \cdots & \sigma^2_{X_n} \end{pmatrix}$$

Les valeurs $\sigma^2_{X_i}$ sont les variances.
Les valeurs $\sigma_{X_iX_j}$ $(i \neq j)$ sont les covariances.

Matrice de covariance

- La matrice de covariance possède certaines propriétés :
  - la matrice Σ est carrée et symétrique

$$\sigma_{X_i X_j} = \sigma_{X_j X_i} \qquad \Longrightarrow \qquad \Sigma' = \Sigma$$

  - la matrice Σ est définie semi-positive

$$\mathbf{a}'\Sigma\mathbf{a} \geq 0 \quad \forall \mathbf{a} \in \mathbb{R}^n$$

  - pour des variables mutuellement indépendantes, on a

$$X_1 \perp \cdots \perp X_n \quad \Longrightarrow \quad \Sigma = \begin{pmatrix} \sigma^2_{X_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{X_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{X_n} \end{pmatrix}$$

Matrice de corrélation

- La matrice de corrélation $\mathbf{R}$ est la version normalisée de Σ.

$$\mathbf{R} = Corr[\mathbf{X}] = \begin{pmatrix} 1 & \rho_{X_1X_2} & \cdots & \rho_{X_1X_n} \\ \rho_{X_2X_1} & 1 & \cdots & \rho_{X_2X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{X_nX_1} & \rho_{X_nX_2} & \cdots & 1 \end{pmatrix}$$

- On passe facilement de $\mathbf{R}$ à Σ et vice-versa :

$$\Sigma = \mathbf{SRS}$$

$$\mathbf{R} = \mathbf{S}^{-1}\Sigma\mathbf{S}^{-1}$$

$$\text{avec } \mathbf{S} = \begin{pmatrix} \sigma_{X_1} & 0 & \cdots & 0 \\ 0 & \sigma_{X_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{X_n} \end{pmatrix}$$

Espérances conditionnelles

Espérances conditionnelles

$$\mu_{X_i|\mathbf{x}_b} = E[X_i|\mathbf{x}_b] = \begin{cases} \displaystyle\sum_{x_i} x_i p(x_i|\mathbf{x}_b) \\ \\ \displaystyle\int_{-\infty}^{+\infty} x_i f(x_i|\mathbf{x}_b) dx_i \end{cases}$$

Si $\mathbf{X}_a = (X_1, \ldots, X_{n_a})'$, on aura

$$\boldsymbol{\mu}_{a|b} = E[\mathbf{X}_a|\mathbf{x}_b] = \begin{pmatrix} \mu_{X_1|\mathbf{x}_b} \\ \vdots \\ \mu_{X_{n_a}|\mathbf{x}_b} \end{pmatrix}$$

- Variances conditionnelles

$$Var[X_i|\mathbf{x}_b] = \begin{cases} \left(\sum_{x_i} x_i^2 p(x_i|\mathbf{x}_b)\right) - \mu^2_{X_i|\mathbf{x}_b} \\ \\ \int_{-\infty}^{+\infty} x_i^2 f(x_i|\mathbf{x}_b) dx_i - \mu^2_{X_i|\mathbf{x}_b} \end{cases}$$

- Covariances conditionnelles

$$Cov[X_i, X_j|\mathbf{x}_b] = \begin{cases} \left(\sum_{x_i}\sum_{x_j} x_i x_j p(x_i, x_j|\mathbf{x}_b)\right) - \mu_{X_i|\mathbf{x}_b}\mu_{X_j|\mathbf{x}_b} \\ \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_i x_j f(x_i, x_j|\mathbf{x}_b) dx_i dx_j - \mu_{X_i|\mathbf{x}_b}\mu_{X_j|\mathbf{x}_b} \end{cases}$$

Matrice de covariance conditionnelle

- Si $\mathbf{X}_a = (X_1, \ldots, X_{n_a})'$, on aura

$$\Sigma_{a|b} = Cov[\mathbf{X}_a|\mathbf{x}_b] = \begin{pmatrix} \sigma^2_{X_1|\mathbf{x}_b} & \sigma_{X_1X_2|\mathbf{x}_b} & \cdots & \sigma_{X_1X_{n_a}|\mathbf{x}_b} \\ \sigma_{X_2X_1|\mathbf{x}_b} & \sigma^2_{X_2|\mathbf{x}_b} & \cdots & \sigma_{X_2X_{n_a}|\mathbf{x}_b} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_{n_a}X_1|\mathbf{x}_b} & \sigma_{X_{n_a}X_2|\mathbf{x}_b} & \cdots & \sigma^2_{X_{n_a}|\mathbf{x}_b} \end{pmatrix}$$

- coefficients de corrélation conditionnels

$$\rho_{X_iX_j|\mathbf{x}_b} = Corr[X_i, X_j|\mathbf{x}_b] = \frac{Cov[X_i, X_j|\mathbf{x}_b]}{\sqrt{Var[X_i|\mathbf{x}_b]Var[X_j|\mathbf{x}_b]}}$$

- Soient $\mathbf{Y} = \mathbf{h}(\mathbf{X})$ des fonctions linéaires des variables $\mathbf{X}$, avec

$$\begin{cases} Y_1 = h_1(\mathbf{X}) = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n \\ Y_2 = h_2(\mathbf{X}) = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n \\ \quad \vdots \\ Y_n = h_n(\mathbf{X}) = a_{n1}X_1 + a_{n2}X_2 + \cdots + a_{nn}X_n \end{cases}$$

ce que l'on note plus facilement $\mathbf{Y} = \mathbf{A}\mathbf{X}$, avec

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

Cas des fonctions linéaires d'un vecteur aléatoire

- Le vecteur espérance et la matrice de covariance sont donnés par

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b} \quad \Longrightarrow \quad \begin{cases} \boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{A}\boldsymbol{\mu}_{\mathbf{X}} + \mathbf{b} \\ \boldsymbol{\Sigma}_{\mathbf{Y}} = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{A}' \end{cases}$$

- Lorsque le vecteur $\mathbf{Y}$ ne contient qu'une variable Y, on obtient

$$Y = \sum_j a_j X_j + b \quad \Longrightarrow \quad \begin{cases} \mu_Y = \sum_j a_j \mu_{X_j} + b \\ \sigma_Y^2 = \sum_i \sum_j a_i a_j Cov[X_i, X_j] \end{cases}$$

Exemple

- Soit le vecteur $\mathbf{X} = (X_1, X_2, X_3)'$, avec
  - $X_1 \equiv$"température maximale journalière
  - $X_2 \equiv$"température minimale journalière
  - $X_3 \equiv$"température moyenne journalière

$$\boldsymbol{\mu}_{\mathbf{X}} = (15, 5, 10)' \quad ; \quad (\sigma^2_{X_1}, \sigma^2_{X_2}, \sigma^2_{X_3}) = (1, 1, 0.5)'$$

$$\mathbf{R}_{\mathbf{X}} = \begin{pmatrix} 1 & 0.7 & 0.9 \\ 0.7 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix}$$

- On s'intéresse à $\mathbf{Y} = (X_1 - X_2, \dfrac{X_1 + X_2}{2}, X_3 - \dfrac{X_1 + X_2}{2})'$
  - que valent $\boldsymbol{\mu}_{\mathbf{Y}}$ et Σ_Y ?
  - quelle est la corrélation entre X_3 et $(X_1 + X_2)/2$?

Vecteur aléatoire normal $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$

- Soient $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ des variables continues.
  On parlera de vecteur aléatoire normal lorsque

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \qquad \forall \mathbf{x} \in \mathbb{R}^n$$

  où Σ est une matrice définie positive

- Le vecteur aléatoire normal a des propriétés remarquables

Distributions marginales et conditionnelles

- Soit une partition du vecteur aléatoire normal $\mathbf{X}$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_a \\ \mathbf{X}_b \end{pmatrix} \quad \text{avec} \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} \; ; \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

  - Distributions marginales :

$$\mathbf{X}_a \sim N(\boldsymbol{\mu}_a, \Sigma_{aa}) \qquad \mathbf{X}_b \sim N(\boldsymbol{\mu}_b, \Sigma_{bb})$$

  - Distributions conditionnelles :

$$\mathbf{X}_a | \mathbf{x}_b \sim N(\boldsymbol{\mu}_{a|b}, \Sigma_{a|b})$$

$$\mathbf{X}_b | \mathbf{x}_a \sim N(\boldsymbol{\mu}_{b|a}, \Sigma_{b|a})$$

Distributions marginales et conditionnelles

- Vecteur espérance et matrice de covariance conditionnelles

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \Sigma_{ab}\Sigma_{bb}^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b) \qquad \Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$

$$\boldsymbol{\mu}_{b|a} = \boldsymbol{\mu}_b + \Sigma_{ba}\Sigma_{aa}^{-1}(\mathbf{x}_a - \boldsymbol{\mu}_a) \qquad \Sigma_{b|a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}$$

Exemple (suite)

- Soit le vecteur $\mathbf{X} = (X_1, X_2, X_3)'$, avec $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$
  - $X_1 \equiv$"température maximale journalière
  - $X_2 \equiv$"température minimale journalière
  - $X_3 \equiv$"température moyenne journalière

Quelle est la distribution de la température moyenne journalière connaissant les valeurs x_1 et x_2 des températures extrêmes ?

Combinaisons linéaires d'un vecteur normal

- Soit le vecteur $\mathbf{X} = (X_1, \ldots, X_n)'$, avec $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$
  Soit la combinaison linéaire $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$

  La distribution de $\mathbf{Y}$ est celle d'un vecteur normal, avec

$$\left.\begin{array}{l} \mathbf{X} \sim N(\boldsymbol{\mu}_{\mathbf{X}}, \Sigma_{\mathbf{X}}) \\ \\ \mathbf{Y} = \mathbf{AX} + \mathbf{b} \end{array}\right\} \Longrightarrow \mathbf{Y} \sim N(\mathbf{A}\boldsymbol{\mu}_{\mathbf{X}} + \mathbf{b}, \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}')$$

  Toutes les combinaisons linéaires d'un vecteur normal donnent un nouveau vecteur normal

Exemple (suite)

- Soit le vecteur $\mathbf{X} = (X_1, X_2, X_3)'$, avec $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$
  - $X_1 \equiv$"température maximale journalière
  - $X_2 \equiv$"température minimale journalière
  - $X_3 \equiv$"température moyenne journalière

Soit le vecteur $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ avec

$$\mathbf{Y} = (X_1 - X_2, \frac{X_1 + X_2}{2}, X_3 - \frac{X_1 + X_2}{2})'$$

Quelle est la distribution du vecteur $\mathbf{Y}$?

- **Indépendance mutuelle**

$$X_1 \perp \cdots \perp X_n \quad \Longleftrightarrow \quad \Sigma = \mathbf{S}^2 = \begin{pmatrix} \sigma^2_{X_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{X_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{X_n} \end{pmatrix}$$

  - L'indépendance est équivalente à l'absence de corrélation.
  - Cette équivalence est valable seulement pour le vecteur normal !

Indépendance

Soit la partition $\mathbf{X} = \begin{pmatrix} \mathbf{X}_a \\ \mathbf{X}_b \end{pmatrix} \sim N(\boldsymbol{\mu}, \Sigma)$ avec $\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$

- indépendances partielles :

$$(1) \quad X_1 \perp \cdots \perp X_{n_a} \iff \Sigma_{aa} = \begin{pmatrix} \sigma^2_{X_1} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{X_{n_a}} \end{pmatrix}$$

$$(2) \quad \mathbf{X}_a \perp \mathbf{X}_b \iff \Sigma_{ab} = \Sigma'_{ba} = \mathbf{0}$$

- indépendance conditionnelle :

$$(X_1 \perp \cdots \perp X_{n_a}) | \mathbf{x}_b \iff \Sigma_{a|b} = \begin{pmatrix} \sigma^2_{X_1|\mathbf{x}_b} & 0 & \cdots & 0 \\ 0 & \sigma^2_{X_2|\mathbf{x}_b} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{X_{n_a}|\mathbf{x}_b} \end{pmatrix}$$

8. Statistique - Introduction

Définitions

- population : ensemble (fini ou infini) des individus distincts auxquels on s'intéresse.
- échantillon : sous-ensemble d'individus issus de la population. (échantillon aléatoire = individus pris au hasard)
- inférence : processus consistant à généraliser à la population des résultats obtenus sur base de l'échantillon.

La Statistique propose des méthodes permettant d'inférer, c'est-à-dire de tirer des conclusions sur la population à partir de l'échantillon.

Exemple

On s'intéresse à la taille X des hommes en Belgique, où N est le nombre d'hommes belges vivant à l'instant considéré. Les tailles observables de ces hommes sont $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$. On constitue un échantillon comprenant n hommes pris au hasard (avec $n << N$).

En traduisant ceci, on a donc

- X est la variable aléatoire "taille d'un homme belge"
- $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ est un échantillon aléatoire, avec $X_i \equiv$"taille du $i$ème homme belge mesuré"
- $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ sont les valeurs obtenues après avoir mesuré la taille X_i de chacun des n hommes

Echantillon i.i.d. (indépendant identiquement distribué)

Soit X la variable aléatoire qui nous intéresse.
Soit $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ un échantillon aléatoire.

L'échantillon $\mathbf{X}$ est dit i.i.d. si :

$$\begin{cases} X_1 \sim \cdots \sim X_n & \Longleftrightarrow \quad F_{X_1}(x) = \cdots = F_{X_n}(x) \\ X_1 \perp \cdots \perp X_n & \Longleftrightarrow \quad F(\mathbf{x}) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \end{cases}$$

En pratique, la distribution de X dépend de paramètres $\boldsymbol{\theta}$ inconnus :

$$X \sim F_X(x; \boldsymbol{\theta})$$

Exemple (suite)

Soit $X \equiv$"taille d'un homme belge", avec $X \sim N(\mu, \sigma^2)$
Si l'échantillon $\mathbf{X}$ est i.i.d., alors

$$X_1 \sim X_2 \sim \cdots \sim X_n \Longleftrightarrow X_i \sim N(\mu, \sigma^2)$$

$$X_1 \perp X_2 \perp \cdots \perp X_n \Longleftrightarrow \mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$$

$$\text{avec} \quad \boldsymbol{\mu} = (\mu, \mu, \ldots, \mu)' \quad ; \quad \Sigma = \sigma^2 \mathbf{I}$$

Question : Comment estimer les paramètres μ, σ^2 à partir de $\mathbf{X}$?

Statistique T

- Statistique : fonction observable qui ne dépend que de l'échantillon.

$$T = h(X_1, \ldots, X_n)$$

  - la statistique T est une variable aléatoire
  - elle dépend aussi des paramètres $\boldsymbol{\theta}$ dont dépendent les X_i.
  - exemples :

    – le minimum : $L = \min(X_1, \ldots, X_n) = X_{(1)}$

    – le maximum : $U = \max(X_1, \ldots, X_n) = X_{(n)}$

    – l'étendue : $R = U - L = X_{(n)} - X_{(1)}$

    – la moyenne exp. : $\overline{X} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i$

    – la variance exp. : $S^2 = \dfrac{1}{(n-1)} \displaystyle\sum_{i=1}^{n} (X_i - \overline{X})^2$

Exemple (suite)

On s'intéresse à la taille moyenne μ d'un homme belge.
On dispose d'un échantillon $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ i.i.d.
On suppose que $X_i \sim N(\mu, \sigma^2)$ $\forall i = 1, \ldots, n$.

Pour la moyenne expérimentale, on a :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \Longrightarrow \quad \overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

On dira que la statistique $\overline{X}$ est un estimateur du paramètre μ :

- ☐ en moyenne, on a $E[\overline{X}] = \mu$
- ☐ la valeur $\overline{x}$ tendra vers μ lorsque n augmente.

Estimateur

- Estimateur : statistique T dont la valeur réalisée t donne une "bonne" estimation d'un paramètre θ.

  On notera $T = \widehat{\Theta}$ et $t = \widehat{\theta}$.

- *Question : qu'entend-on par une "bonne" estimation ?*

  - **non biaisée :** en moyenne, la valeur réalisée $\widehat{\theta}$ sera égale à θ

  $$E[\widehat{\Theta}] = \theta \quad \iff \quad b(\widehat{\Theta}) = E[\widehat{\Theta}] - \theta = 0$$

  - **consistante :** $\widehat{\theta}$ est d'autant plus proche de θ que n augmente

  $$\lim_{n \to \infty} P(|\widehat{\Theta} - \theta| < \varepsilon) = 1$$

  ce qui sera vérifié si $\lim_{n \to \infty} Var[\widehat{\Theta}] = 0$

Exemples de "bons" estimateurs ponctuels

Estimateur	Statistique
$\widehat{E}[X^k]$	$\frac{1}{n}\sum_{i=1}^{n} X_i^k$
$\widehat{\mu}$	$\overline{X} = \widehat{E}[X]$
$\widehat{\sigma}^2$	$S^2 = \frac{n}{n-1}\widehat{E}[(X-\mu)^2]$
$\widehat{\sigma}$	$S = \sqrt{S^2}$
$\widehat{F}(x)$	$\frac{\#\{X_i \leq x\}}{n}$
$\widehat{p}(x)$	$\frac{\#\{X_i = x\}}{n}$

Exercice

La moyenne expérimentale est-elle un bon estimateur de la moyenne théorique pour un échantillon i.i.d. ?

- Si $E[X] = \mu, Var[X] = \sigma^2$, alors $E[\overline{X}] = \mu$ et $Var[\overline{X}] = \dfrac{\sigma^2}{n}$
- On a donc :

$$E[\overline{X}] - \mu = 0 \quad \Longrightarrow \quad \overline{X} \text{ est un estimateur non biaisé}$$

$$\lim_{n\to\infty} Var[\overline{X}] = 0 \quad \Longrightarrow \quad \overline{X} \text{ est un estimateur consistant}$$

Conclusion : la statistique $\overline{X}$ est un bon estimateur de μ.

Distribution des estimateurs

- Deux estimateurs ponctuels important sont $\overline{X}$ et S^2.
  Ce sont les estimateurs classiques de μ et σ^2 :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad ; \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

  Puisque ce sont des statistiques, ils ont une distribution.
  On peut montrer que :

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad ; \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \quad ; \quad \overline{X} \perp S^2$$

  Question : comment établir un intervalle de confiance pour μ et σ^2 ?

intervalle de confiance à la moyenne (σ^2 connu)

- On sait que $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- On peut donc écrire que

$$P\left(z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

- En résolvant cette double inégalité par rapport à μ, on obtient

$$P\left(\overline{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n} \leq \mu \leq \overline{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}\right) = 1 - \alpha$$

- Un intervalle de confiance de niveau $1 - \alpha$ est alors donnée par

$$\overline{x} - z_{1-\alpha/2}\sqrt{\sigma^2/n} \leq \mu \leq \overline{x} + z_{1-\alpha/2}\sqrt{\sigma^2/n}$$

Exercice

- Soit X la taille (en cm) des étudiants masculins BIR 12/SINF12. On suppose $X \sim N(\mu, \sigma^2)$, avec μ, σ^2 inconnues.
- On considère un échantillon i.i.d. $\mathbf{X} = (X_1, X_2, \dots, X_{10})'$ de 10 étudiants choisis au hasard en BIR 12/SINF12.
- Les valeurs des mesures de taille $\mathbf{x} = (x_1, x_2, \dots, x_{10})'$ sont :

$$\mathbf{x} = (\quad ,\quad ,\quad ,\quad ,\quad ,\quad ,\quad ,\quad ,\quad ,\quad)'$$

- Sur base de ce qui précède, estimez :
  - ☐ les valeurs μ et σ^2 avec $\overline{x}$ et s^2
  - ☐ la valeur μ avec un intervalle de confiance (remplacez σ^2 par s^2)

Obtention d'un estimateur ponctuel

On a présenté les formules pour quelques "bons" estimateurs ponctuels. Comment les a-t-on obtenues ?

Trois méthodes classiques sont :

- la méthode des moments
- la méthode du maximum de vraisemblance
- la méthode des moindres carrés

Seule la méthode des moindres carré sera illustrée ici, dans le cadre de la régression linéaire.

On dispose de valeurs observées $(\mathbf{y}|\mathbf{x})' = (y|x_1, \ldots, y|x_n)$
On considère un modèle de régression $E[Y|x] = \mu_{Y|x} = h(x; \boldsymbol{\theta})$

Question : Comment obtenir un estimation des paramètres $\boldsymbol{\theta}$?
Réponse : Assez simple si $E[Y|x]$ est une combili des $\boldsymbol{\theta}$, avec

$$\mu_{Y|x} = \theta_1 h_1(x) + \cdots + \theta_p h_p(x)$$

On peut alors écrire

$$\begin{cases} \mu_{Y|x_1} & = \theta_1 h_1(x_1) + \cdots + \theta_p h_p(x_1) \\ \mu_{Y|x_2} & = \theta_1 h_1(x_2) + \cdots + \theta_p h_p(x_2) \\ & \vdots \\ \mu_{Y|x_n} & = \theta_1 h_1(x_n) + \cdots + \theta_p h_p(x_n) \end{cases}$$

Régression - méthode des moindres carrés

- En notation matricielle, on a alors

$$\begin{pmatrix} \mu_{Y|x_1} \\ \vdots \\ \mu_{Y|x_n} \end{pmatrix} = \begin{pmatrix} h_1(x_1) & \cdots & h_p(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_n) & \cdots & h_p(x_n) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} \quad \Longleftrightarrow \quad \boldsymbol{\mu}_{Y|\mathbf{x}} = \mathbf{H}\boldsymbol{\theta}$$

- La somme des carrés d'écarts est donnée par

$$\begin{aligned} SCE(\boldsymbol{\theta}; \mathbf{y}|\mathbf{x}) &= \sum_{i=1}^{n} \left(y|x_i - \mu_{Y|x_i}\right)^2 \\ &= (\mathbf{y}|\mathbf{x} - \boldsymbol{\mu}_{Y|\mathbf{x}})'(\mathbf{y}|\mathbf{x} - \boldsymbol{\mu}_{Y|\mathbf{x}}) \\ &= (\mathbf{y}|\mathbf{x} - \mathbf{H}\boldsymbol{\theta})'(\mathbf{y}|\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \\ &= (\mathbf{y}|\mathbf{x})'(\mathbf{y}|\mathbf{x}) - 2(\mathbf{y}|\mathbf{x})'\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}'\mathbf{H}'\mathbf{H}\boldsymbol{\theta} \end{aligned}$$

Régression - méthode des moindres carrés

- $SCE(\boldsymbol{\theta};\mathbf{y}|\mathbf{x})$ mesure l'écart entre valeurs attendues et observées
  On choisira comme paramètres $\widehat{\boldsymbol{\theta}}$ ceux pour lesquels

$$SCE(\widehat{\boldsymbol{\theta}};\mathbf{x}) = \arg\min_{\theta} SCE(\boldsymbol{\theta};\mathbf{x})$$

- Lorsque la fonction est convexe par rapport aux $\boldsymbol{\theta}$,

$$\left.\frac{\partial SCE(\boldsymbol{\theta};\mathbf{x})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \mathbf{0}$$

$\Longrightarrow$ on a comme résultat que

$$\left.\frac{\partial SCE(\boldsymbol{\theta};\mathbf{x})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -2\mathbf{H}'(\mathbf{y}|\mathbf{x}) + 2\mathbf{H}'\mathbf{H}\widehat{\boldsymbol{\theta}} = \mathbf{0}$$

$$\Longleftrightarrow \qquad \widehat{\boldsymbol{\theta}} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'(\mathbf{y}|\mathbf{x})$$

Exemple

Afin de lier le rendement $Y|x$ d'une culture de blé (en quintal par hectare) à la dose x d'engrais azoté appliquée sur cette culture (en kilogramme par hectare), on a choisi cinq champs sur lesquels on a appliqué différentes doses de cet engrais.

Les résutats sont les suivants :

x_i	1	2	3	4	5
$y\|x_i$	60	71	75	87	99

On envisage le modèle $\mu_{Y|x} = \theta_1 + \theta_2 x$.
Quelles sont les valeurs estimées des paramètres par moindres carrés ?

Exemple (suite)

- En notation matricielle, on a donc

$$\begin{pmatrix} \mu_{Y|x_1} \\ \mu_{Y|x_2} \\ \mu_{Y|x_3} \\ \mu_{Y|x_4} \\ \mu_{Y|x_5} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \quad ; \quad \begin{pmatrix} y|x_1 \\ y|x_2 \\ y|x_3 \\ y|x_4 \\ y|x_5 \end{pmatrix} = \begin{pmatrix} 60 \\ 71 \\ 75 \\ 87 \\ 99 \end{pmatrix}$$

$$\Longrightarrow \quad \widehat{\boldsymbol{\theta}} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'(\mathbf{y}|\mathbf{x}) = \begin{pmatrix} 50.2 \\ 9.4 \end{pmatrix}$$