Master IAD

Module MI009 – RFIDEC

Transparents de cours

année 2012–2013

Christophe Gonzales

RFIDEC — cours 1 :
Rappels de probas/stats (1/3)

Christophe Gonzales

LIP6 – Université Paris 6, France

Modalités de contrôle

- Pas de contrôle continu
- Note finale d'UE en trois parties :
  1. Examen écrit en milieu de semestre (35%)
  2. Examen écrit en fin de semestre (35%)
  3. Contrôle sur machine (30%)

Introduction à la reconnaissance des formes

Reconnaissance des formes

Algorithmes $\Longrightarrow$ reconnaissance automatique de régularités, de motifs dans un amas de données.

Utilisation de ces motifs pour des tâches (classification, etc)

Exemple :

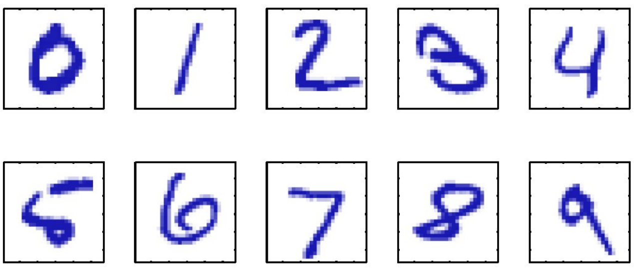
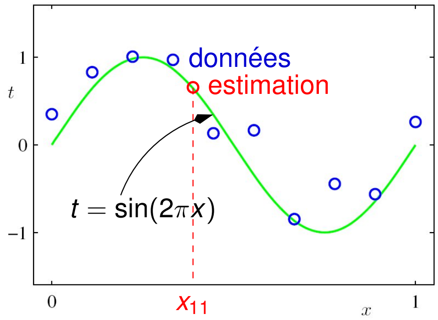


Image = données (pixels) *Classification* = reconnaître le chiffre

Exemple : problème d'ajustement (1/5)



Observations
(x_1, t_1) $\vdots$ (x_{10}, t_{10})

$\Longrightarrow$ courbe $\sin(2\pi x)$ $\Longrightarrow$ estimation de t_{11}

$\Longrightarrow$ reconnaissance de la courbe verte

Exemple : problème d'ajustement (2/5)

Idée : estimer la courbe verte par un polynôme :

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

$\Longrightarrow$ Problèmes :

- déterminer les « meilleurs » coefficients w_j
- déterminer la « meilleure » valeur de M

« meilleur » $\Longrightarrow$ critère d'optimalité

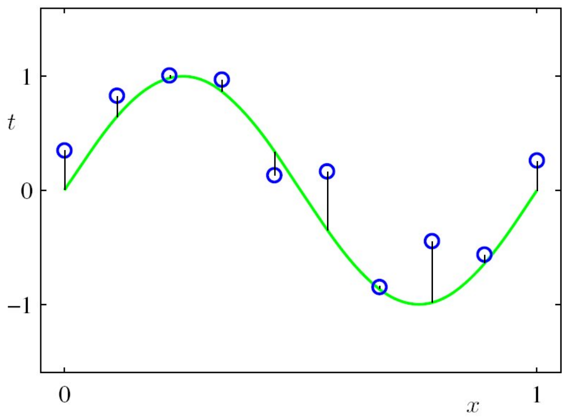
critère d'optimalité possible

Minimiser une fonction d'erreur mesurant l'inadéquation entre la courbe $y(x, \mathbf{w})$ et les points observés :

$$E(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{10}[y(x_i, \mathbf{w}) - t_i]^2$$

Exemple : problème d'ajustement (3/5)

$$E(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{10}[y(x_i, \mathbf{w}) - t_i]^2$$



Exemple : problème d'ajustement (4/5)

Problème 1 : $\mathbf{w}^*$ tel que $\mathbf{w}^* = \min_{\mathbf{w}} E(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{10}[y(x_i,\mathbf{w}) - t_i]^2$

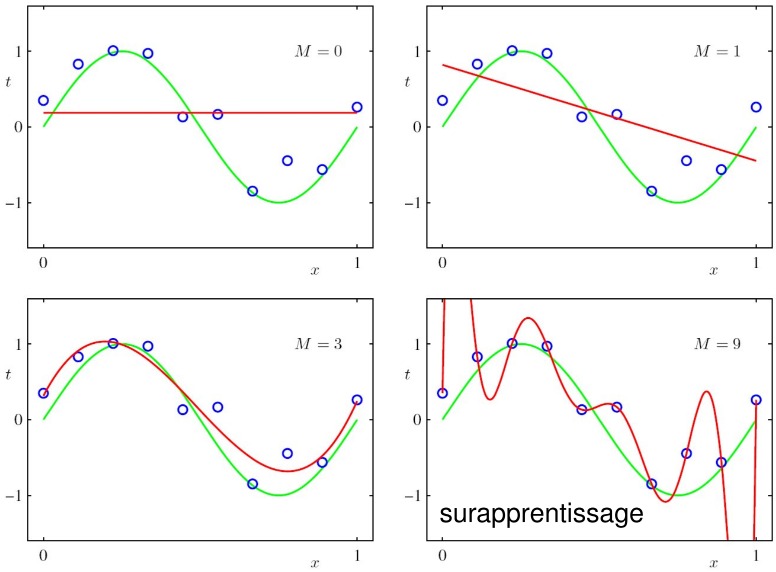
$E(\mathbf{w})$ fonction quadratique en $\mathbf{w}$ $\Longrightarrow$ 1 seule solution

$\Longrightarrow \frac{dE(\mathbf{w})}{d\mathbf{w}} = 0$

$\Longrightarrow$ résoudre un système linéaire

Problème 2 : Choix de M $\Longrightarrow$ sélection de modèle

Exemple : problème d'ajustement (5/5)



Plan général du cours

1. Introduction et rappels de proba/stat
2. Estimations ponctuelles à partir d'échantillons
3. Intervalles de confiance, tests d'hypothèses
4. Tests d'ajustement et maximum de vraisemblance
5. MAP et apprentissage non paramétrique
6. Régression linéaire
7. Clustering (K-means, Kohonen)
8. Classification bayésienne
9. Classification gaussienne
10. Classifieur linéaire : LDA, perceptron

Plan du cours 1.1

1. Introduction à la statistique descriptive
2. vocabulaire de stat descriptive
3. distributions et représentations
4. indicateurs de « moyenne »
5. indicateurs de dispersion

Intro à la statistique descriptive

Statistique

- recueil de données
- traitement de ces données
- l'exploitation (interprétation, prévision, etc) sort du cadre de la stat descriptive

Principe de la statistique descriptive

- des amas de données
- stat descriptive $\Longrightarrow$ synthèse, résumé d'informations

données synthétiques $\Longrightarrow$ exploitation

⚠ pas forcément besoin de probabilités

Exemple

Salaires de cadres masculins (exprimés en k€)

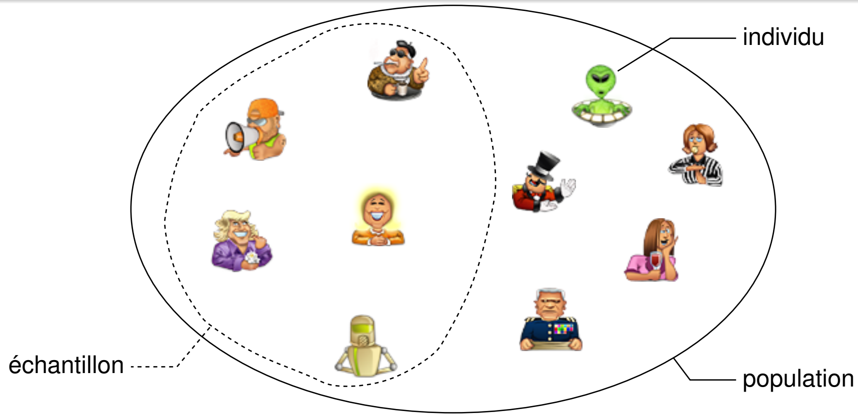
18	10	12	12	10	14	14	14	12	12	16	18	10
16	16	12	12	14	14	14	14	12	16	14	14	16
12	10	18	12	18	18	14	14	14	12	16	14	16
10	16	16	14	14	14	12	16	16	14	14	12	14

Salaires de cadres féminins (exprimés en k€)

16	16	14	14	14	14	10	12	12	08	20	14	14
14	20	14	12	12	16	10	10	18	10	16	12	12
14	18	12	12	16	08	14	10	12	14	12	16	14

Question : Les hommes sont-ils mieux payés que les femmes ?

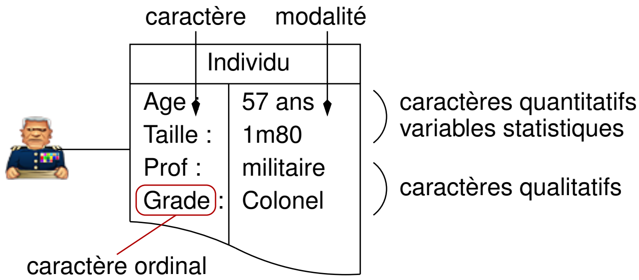
Vocabulaire (1/3)



Définitions

- population (statistique) : ensemble des objets (ou personnes) sur lesquels porte l'étude
- individu : chaque élément de la population

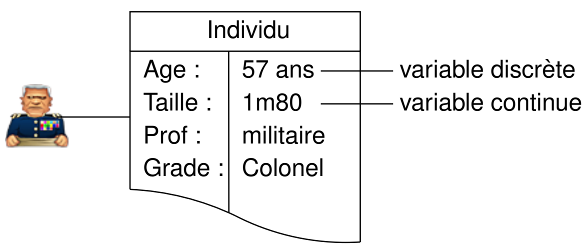
Vocabulaire (2/3)



Définitions

- Caractères : critères d'étude de la population
- Modalités : les valeurs que peuvent prendre les caractères
- Caractère quantitatif ou Variable statistique : ensemble de modalités $=$ des nombres $+$ échelle mathématique
- Caractère qualitatif ou Variable catégorielle : caractère non quantitatif
- Caractère ordinal : les modalités sont ordonnées

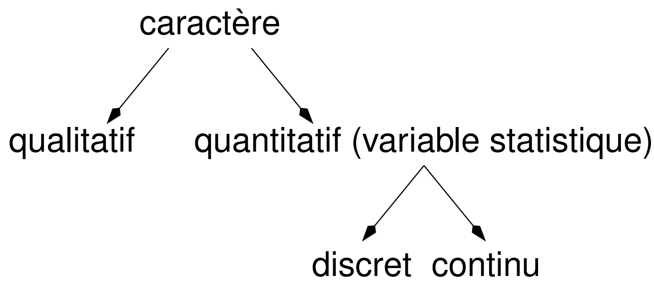
Vocabulaire (3/3)



Définitions sur les variables statistiques

- Variable discrète : définie sur un espace discret (par exemple des entiers)
- Variable continue : définie sur un continuum (toutes les valeurs numériques d'un intervalle)

Résumé : classification des caractères



Effectifs, fréquences et distributions

Quelques définitions

- X : caractère défini sur une population de N individus
- $\{x_1, \dots, x_I\}$ modalités de X
- N_i = effectif de x_i
  = nombre d'individus pour lesquels X a pris la valeur x_i
- fréquence ou effectif relatif : $f_i = \frac{N_i}{N}$
- distribution de X : ensemble des couples $\{(x_1, f_1), (x_2, f_2), \dots\}$

Idée :

calculer la distribution des caractères sur l'amas de données
$\Longrightarrow$ résume les informations importantes

Diagrammes en bâtons

Définition du diagramme en bâtons

- caractère X de distribution $\{(x_1, f_1), (x_2, f_2), (x_3, f_3), \dots\}$
- diagramme en bâtons = graphe dans lequel on associe à chacune des modalités x_i (représentées sur l'axe horizontal) un bâton de hauteur f_i
- représentation en colonnes : idem mais en élargissant les bâtons

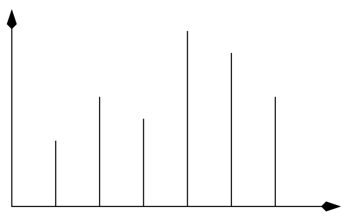
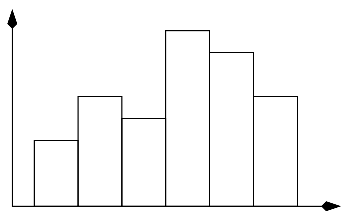
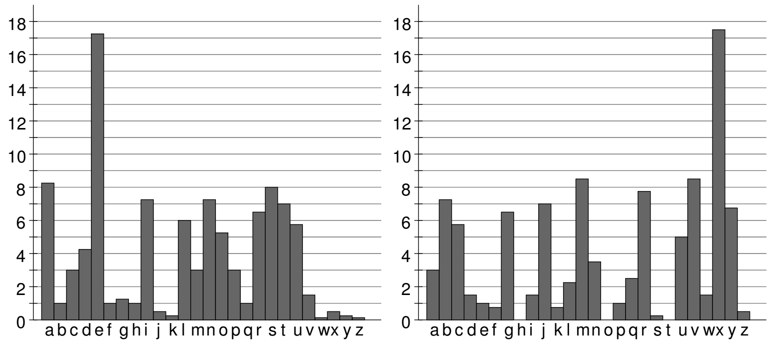


diagramme en bâtons


représentation en colonnes

Application du diagramme en bâtons

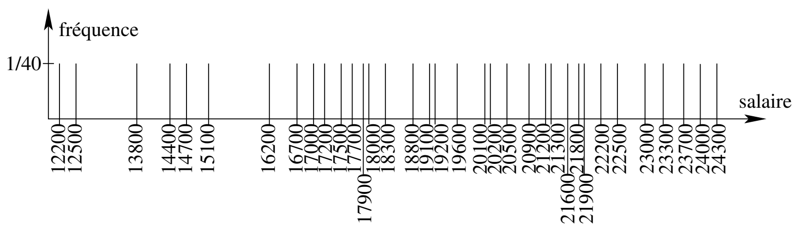
```
XY AXJ BYRJMYJ, MQQMVUVXYJ GXR NCBWJR N'UYX
LMBY N'PCLLX XJ BGR XAVBDBVXYJ, XY IMAX NU
AMYNXGMFVX, RUV GX QGMJVX NU LUV NU QMGMBR
VCEMG.
```



fréquence théorique des lettres en français — fréquence des lettres du cryptogramme

Problèmes des diagrammes en bâtons

Salaires des cadres masculins (en k€)								
19,6	12,5	17,7	18,8	19,1	14,7	21,9	22,5	21,8
20,1	16,2	20,5	12,2	25,4	20,9	21,2	15,5	21,3
17,9	25,0	14,4	21,7	18,3	16,7	23,0	17,0	24,3
27,0	27,7	18,0	17,5	23,7	21,6	13,8	22,2	19,2
20,2	15,1	19,6	17,2	23,3	24,0			



variables continues $\Longrightarrow$ diagramme en bâtons inutilisable

Discrétisation

Idée force : utiliser les diagrammes en bâton ssi la taille de la population $\gg$ au nombre de modalités des caractères

sinon $\Longrightarrow$ discrétiser le caractère

Discrétisation

regrouper toutes les modalités appartenant à certains intervalles dans des classes de données

Caractéristiques d'une classe

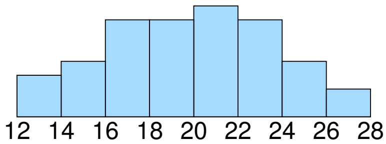
- $\mathcal{C}$: la classe $[a, b[$
- centre de la classe $\mathcal{C}$: $c = \frac{a+b}{2}$
- amplitude de la classe $\mathcal{C}$: $h = b - a$

Exemple de discrétisation (1/3)

Salaires des cadres masculins (en k€)								
19,6	12,5	17,7	18,8	19,1	14,7	21,9	22,5	21,8
20,1	16,2	20,5	12,2	25,4	20,9	21,2	15,5	21,3
17,9	25,0	14,4	21,7	18,3	16,7	23,0	17,0	24,3
27,0	27,7	18,0	17,5	23,7	21,6	13,8	22,2	19,2
20,2	15,1	19,6	17,2	23,3	24,0			

Classe de salaire	effectif	fréq.	Classe de salaire	effectif	fréq.
$[12000, 14000[$	3	$3/42$	$[20000, 22000[$	8	$8/42$
$[14000, 16000[$	4	$4/42$	$[22000, 24000[$	7	$7/42$
$[16000, 18000[$	7	$7/42$	$[24000, 26000[$	4	$4/42$
$[18000, 20000[$	7	$7/42$	$[26000, 28000[$	2	$2/42$

Diagramme
en colonnes :

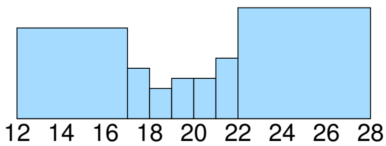


Exemple de discrétisation (2/3)

Salaires des cadres masculins (en k€)								
19,6	12,5	17,7	18,8	19,1	14,7	21,9	22,5	21,8
20,1	16,2	20,5	12,2	25,4	20,9	21,2	15,5	21,3
17,9	25,0	14,4	21,7	18,3	16,7	23,0	17,0	24,3
27,0	27,7	18,0	17,5	23,7	21,6	13,8	22,2	19,2
20,2	15,1	19,6	17,2	23,3	24,0			

Classe de salaire	effectif	fréq.	Classe de salaire	effectif	fréq.
$[12000, 17000[$	9	$9/42$	$[20000, 21000[$	4	$4/42$
$[17000, 18000[$	5	$5/42$	$[21000, 22000[$	6	$6/42$
$[18000, 19000[$	3	$3/42$	$[22000, 28000[$	11	$11/42$
$[19000, 20000[$	4	$4/42$			

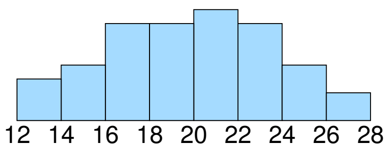
Diagramme
en colonnes :



Exemple de discrétisation (3/3)

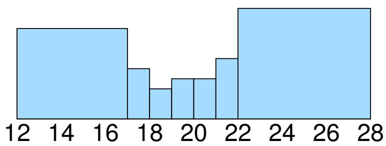
Première discrétisation :

Diagramme
en colonnes :



Deuxième discrétisation :

Diagramme
en colonnes :



N'utiliser le diagramme en colonnes que pour des discrétisations uniformes

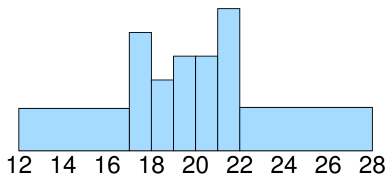
Histogrammes

Définition d'un histogramme

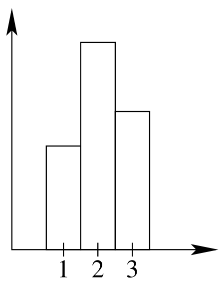
graphe dont l'axe des abscisses représente les modalités et qui associe à chaque classe de modalités un rectangle dont la base correspond aux bornes de cette classe et dont la hauteur est égale à :

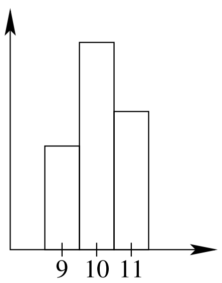
$$\text{hauteur} = \frac{\text{fréquence}}{\text{base}}.$$

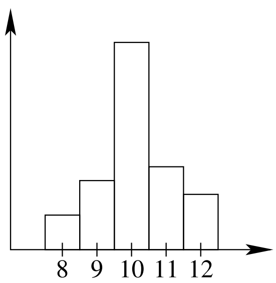
$\Longrightarrow$ la fréquence d'une classe est égale à la surface de son rectangle



Synthèse d'histogrammes ?







Caractéristiques statistiques :

- localisation sur l'axe des $X \Longrightarrow$ moyenne, médiane, mode
- étendue $\Longrightarrow$ écart-type, variance

La moyenne

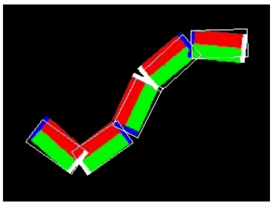
Moyenne d'une variable statistique discrète

- X : variable statistique discrète
- modalités : $\{x_1, x_2, \dots, x_I\}$
- population de N individus
- Moyenne de X : $\mu_X = \sum_{i=1}^{I} f_i x_i = \frac{1}{N} \sum_{i=1}^{I} N_i x_i$

Moyenne d'une variable statistique continue

- Y : variable statistique continue
- modalités regroupées en I classes de centres $\{y_1, y_2, \dots, y_I\}$
- population de N individus
- Moyenne de Y : $\mu_Y = \sum_{i=1}^{I} f_i y_i = \frac{1}{N} \sum_{i=1}^{I} N_i y_i$

Application de la moyenne



Propriétés de la moyenne

Propriété 1

- X : variable statistique
- $Y = aX + b$, où a et b sont des nombres réels
- $\mu_Y = a\mu_X + b$

Propriété 2

- X : variable statistique
- n sous-populations distinctes $S_1, \dots, S_n$ de $N_1, \dots, N_n$ individus
- μ_i la moyenne de X sur la i^ème sous-population
- μ_X : moyenne de X sur la population $S = S_1 \cup S_2 \cup \ldots \cup S_n$

$$\mu_X = \frac{\sum_{i=1}^{n} N_i \mu_i}{\sum_{i=1}^{n} N_i}.$$

Médiane d'une variable statistique discrète

- *moyenne* = valeur centrale des modalités
- *médiane* = valeur de la variable pour laquelle la moitié de la population a une modalité inférieure à cette valeur et l'autre moitié en a une supérieure

Médiane d'une variable statistique discrète

- X : variable statistique discrète, modalités : $\{x_1, \dots, x_l\}$
- population de N individus
- médiane de X = tout nombre δ tel que :

$$\sum_{i \in \{j : x_j < \delta\}} N_i \leq N/2 \quad et \quad \sum_{i \in \{j : x_j > \delta\}} N_i \leq N/2$$

Exemples

variable statistique X						
modalité	0	1	2	3	4	5
effectif	2	3	4	5	3	2

Données triées
0, 0, 1, 1, 1, 2, 2, 2, 2, $\boxed{3}$, 3, 3, 3, 3, 4, 4, 4, 5, 5

---

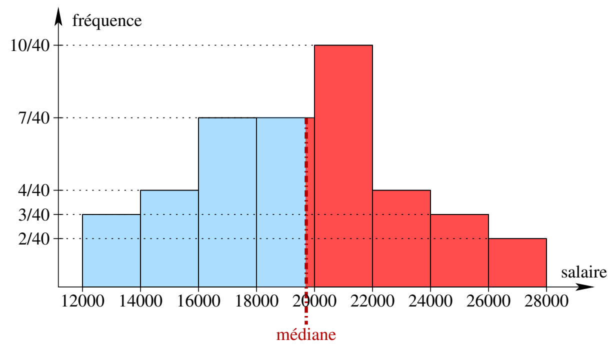
variable statistique X						
modalité	0	1	2	3	4	5
effectif	2	3	4	4	3	2

Données triées
0, 0, 1, 1, 1, 2, 2, 2, $\boxed{2}$, $\boxed{3}$, 3, 3, 3, 4, 4, 4, 5, 5

Médiane d'une variable statistique continue

Médiane d'une variable statistique continue

- X : variable statistique continue
- Médiane $=$ le nombre δ tel que les aires situées de part et d'autre de ce nombre dans l'histogramme représentant X sont égales



Les quantiles

Quantile d'une variable discrète

- X : variable statistique discrète, modalités $\{x_1, \ldots, x_I\}$
- population de N individus
- quantile d'ordre α $=$ tout nombre δ tel que :

$$\sum_{i \in \{j : x_j < \delta\}} N_i \leq \alpha N \quad \text{et} \quad \sum_{i \in \{j : x_j > \delta\}} N_i \leq (1 - \alpha) N$$

Quantile d'une variable continue

- X : variable statistique continue
- quantile d'ordre α $=$ le nombre δ tel que les aires situées de part et d'autre de ce nombre dans l'histogramme représentant X sont égales respectivement à $\alpha \times$ aire totale et $(1 - \alpha) \times$ aire totale

Principaux quantiles

Principaux quantiles	
valeur de α	nom du quantile d'ordre α
$1/2$	médiane
$i/4$ ($i = 1, 2, 3$)	$i^{\text{ème}}$ quartile
$i/5$ ($i = 1, 2, 3, 4$)	$i^{\text{ème}}$ quintile
$i/10$ ($i = 1, 2, \ldots, 9$)	$i^{\text{ème}}$ décile
$i/100$ ($i = 1, \ldots, 99$)	$i^{\text{ème}}$ centile

Propriété des quantiles

Propriété

- X : variable statistique
- $Y = aX + b$, où a et b sont des nombres réels
- $\text{médiane}(Y) = a \times \text{médiane}(X) + b$

Application de la médiane

Filtre médian :

x_{11}	x_{12}	x_{13}
x_{21}	x_{22}	x_{23}
x_{31}	x_{32}	x_{33}

$\longrightarrow$

	M	

$M = \text{médiane des } x_{ij}$

Étendue des données

Étendue d'une variable discrète

- X : variable statistique discrète
- Étendue de X = la différence entre la plus grande modalité de X et la plus petite modalité

Étendue d'une variable continue

- X : variable statistique continue
- Étendue de X = la différence entre la borne supérieure de la classe associée à la plus grande valeur observée et la borne inférieure de la classe associée à la plus petite valeur observée

⚠ Mesure de dispersion peu utilisée

Variance (1/2)

Idée force : analyser globalement les déviations entre les valeurs prises sur un individu et la moyenne de la population

Variance d'une variable discrète

- X : variable statistique discrète, modalités : $x_1, \dots, x_I$
- population : N individus
- Variance de X : $\sigma_X^2 = \sum_{i=1}^{I} f_i(x_i - \mu_X)^2 = \frac{1}{N}\sum_{i=1}^{I} N_i(x_i - \mu_X)^2$

Variance d'une variable continue

- Y : variable statistique continue
- modalités : I classes de centres $\{y_1, y_2, \dots, y_I\}$
- population : N individus
- Variance de Y : $\sigma_Y^2 = \sum_{i=1}^{I} f_i(y_i - \mu_Y)^2 = \frac{1}{N}\sum_{i=1}^{I} N_i(y_i - \mu_Y)^2$

Variance (2/2)

Calcul pratique de la variance

$$\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{I} N_i(x_i - \mu_X)^2$$

$$= \frac{1}{N}\left(\sum_{i=1}^{I} N_i x_i^2\right) - \mu_X^2$$

Propriété

- X : variable statistique
- $Y = aX + b$, où a et b sont des nombres réels
- $\sigma_Y^2 = a^2\sigma_X^2$

Écart-type

Problème de la variance : unités différentes des données

X exprimé en € $\Longrightarrow \sigma_X^2$ exprimé en €^{2}

Écart-type

- X : variable statistique (discrète ou continue)
- Écart-type de X : σ_X, la racine carrée de la variance de X

Propriété

- X : variable statistique
- $Y = aX + b$, où a et b sont des nombres réels
- $\sigma_Y = |a|\sigma_X$

Propriété de l'écart-type

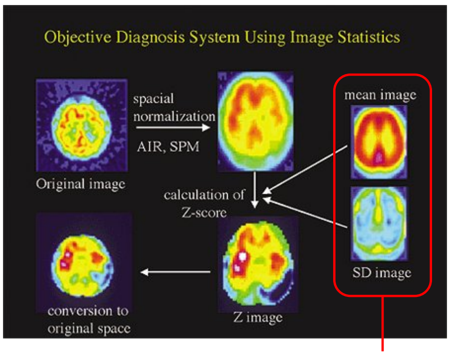
Théorème de Bienaymé-Tchébicheff

- X une variable statistique discrète
- moyenne μ_X et écart-type σ_X
- la proportion de valeurs de X observées entre $\mu_X - k\sigma_X$ et $\mu_X + k\sigma_X$ est strictement supérieure à $1 - 1/k^2$
- Propriété valable pour les variables statistiques continues :

la proportion d'individus pour lesquels la valeur de X se situe entre $\mu_X - k\sigma_X$ et $\mu_X + k\sigma_X$ est strictement supérieure à $1 - 1/k^2$

$$\Longrightarrow \begin{cases} \text{plus de 88\% de la population entre } \mu_X - 3\sigma_X \text{ et } \mu_X + 3\sigma_X \\ \text{plus de 96\% de la population entre } \mu_X - 5\sigma_X \text{ et } \mu_X + 5\sigma_X \end{cases}$$

Application de l'écart-type



base de 1200 patients

1. normaliser l'image du cerveau du patient (SPM : Statistical Parametric Mapping)
2. comparer avec la base de patients
3. produire l'image comparée
4. diagnostiquer la maladie

RFIDEC — cours 1 :
Rappels de probas/stats (2/3)

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours 1.2

1. probabilités : événements, définition
2. probabilités conditionnelles
3. formule de Bayes
4. indépendance, indépendance conditionnelle

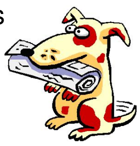
Les probabilités : objectives ou subjectives ? (1/2)

Qu'est-ce qu'une probabilité ?

Probabilités subjectives (de Finetti)

Les probabilités expriment le degré d'incertitude que les individus accordent à des événements incertains.

Exemple : dans un journal, « certains observateurs de la vie politique pensent qu'il y a une chance sur quatre pour que la rencontre au sommet ait lieu avant novembre »



$\Longrightarrow P(\text{rencontre avant novembre}) = 0,25$

Les probabilités : objectives ou subjectives ? (2/2)

Probabilités objectives

Les probabilités ne dépendent pas d'appréciations personnelles et doivent être fondées uniquement sur des données *objectives*.

$\Longrightarrow$ probas introduites au XVIème, XVIIème et XVIIIème siècles (Cardano, De Moivre, Pascal, Bernoulli, Laplace)

$\Longrightarrow$ données fréquentistes

$\Longrightarrow$ dans les années 30 : axiomatique de Kolmogorov

Les probabilités : outil indispensable

Richard T. Cox

- machine pour gérer des incertitudes
- Rationalité :
  $\Longrightarrow$ probas = seule représentation possible des incertitudes

rationnel = 5 desideratas :

1. incertitude d'un événement évaluée de plusieurs manières $\Longrightarrow$ même résultat
2. petites modif d'un événement $\Longrightarrow$ petit changement sur l'incertitude
3. représentation universelle : s'applique à tout problème
4. évaluer l'incertitude d'un événement $\Longrightarrow$ événement défini sans ambiguïté
5. évaluation $\Longrightarrow$ utiliser toutes les infos à disposition

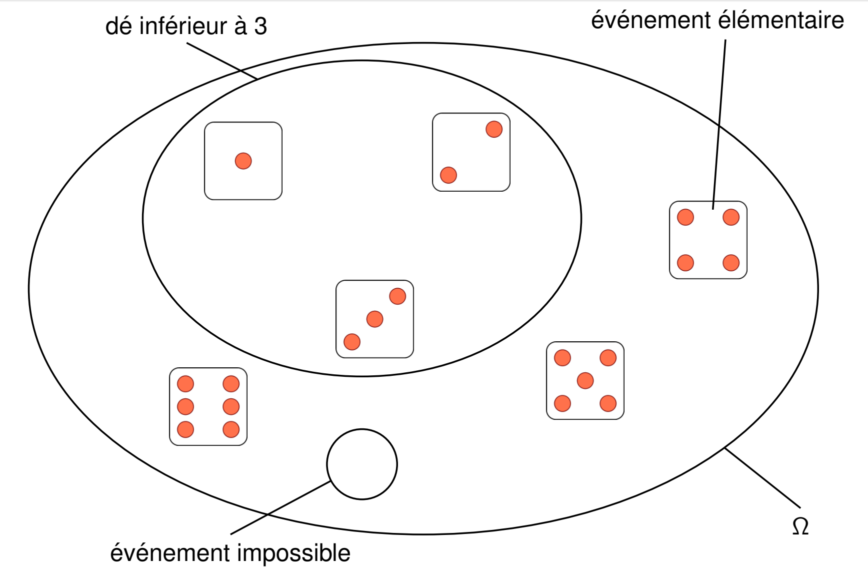
Vocabulaire

Exemple : lancer d'un dé à 6 faces

Définition

- événement : tout ce qui peut se réaliser ou pas à la suite d'une expérience
  exemple : « obtenir un 4 », « ne pas obtenir un 4 », « obtenir un chiffre inférieur à 3 »
- événement certain : assuré de se produire
  exemple : « obtenir un chiffre inférieur à 7 »
- événement impossible : ne se produira jamais
  exemple : « obtenir un chiffre supérieur à 7 »
- événement élémentaire : seulement un seul résultat de l'expérience permet de le réaliser
  exemple : « obtenir 4 » mais « obtenir un chiffre pair » = pas élémentaire
- univers Ω : ensemble des événements élémentaires

Vocabulaire

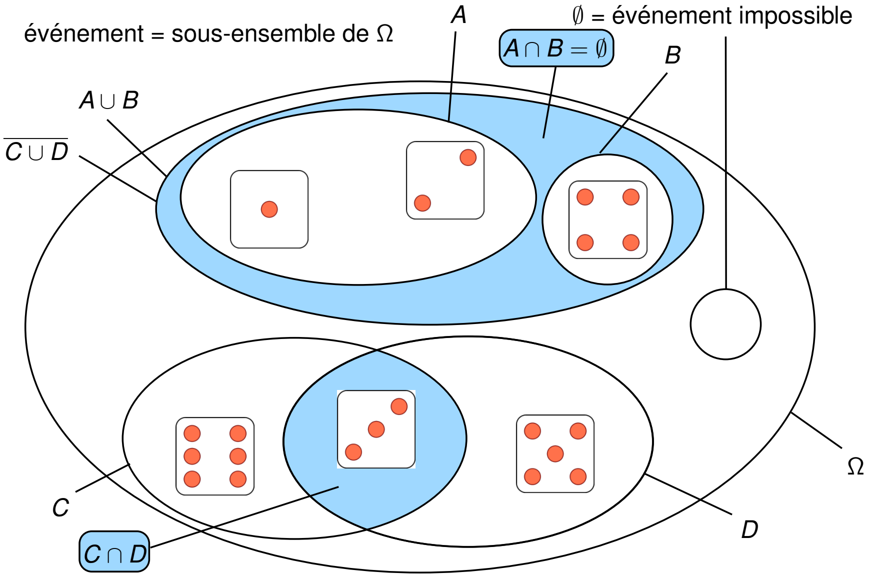


Les événements, des ensembles ?

- événements = sous-ensembles de Ω
- $\emptyset$ = événement impossible
- $A \cup B$ = événement qui est réalisé si A ou B est réalisé
- $C \cap D$ = événement qui est réalisé si C et D sont réalisés
- $\overline{C \cup D}$ = complémentaire de $C \cup D$ dans Ω

  = événement qui est réalisé ssi $C \cup D$ ne l'est pas
- $A \cap B = \emptyset$ = 2 événements qui ne peuvent se réaliser simultanément

Les événements, des ensembles ?



Définition des probabilités : le cas discret

Définition des probabilités (Kolmogorov)

- Ω = ensemble fini ou dénombrable d'événements élémentaires e_k, $k \in K \subseteq \mathbb{N}$
- $\mathcal{A} = 2^\Omega$ = ensemble des événements
- pour tout $A \in \mathcal{A}$, $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- $A = \bigcup_{k \in L} A_k$, avec L ensemble dénombrable et, $\forall j, k \in L$, $j \neq k$, $A_j \cap A_k = \emptyset$, $P(A) = \sum_{k \in L} P(A_k)$.

$\Longrightarrow$ Les probabilités des événements élémentaires déterminent entièrement P

conséquence : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Définition des probabilités : le cas continu (1/3)

Chaque événement élémentaire a une proba = 0

Mais proba d'être dans un intervalle $\neq 0$

1ère étape : discrétisation

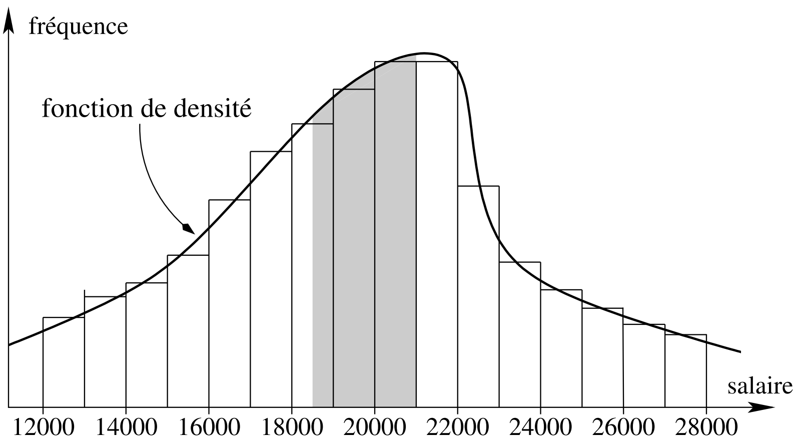
transformer Ω en un ensemble dénombrable de classes

Exemple : salaires $\Longrightarrow$ 8 classes :

[12K€, 14K€[, [14K€, 16K€[, [16K€, 18K€[, ... , [26K€, $+\infty$[

Définition des probabilités : le cas continu (2/3)

Histogramme : surface = fréquence $\approx$ proba



Définition des probabilités : le cas continu (3/3)

Probabilité d'un événement

- fonction de densité
- $P(A)$ = surface délimitée par la fonction de densité dans la zone où les événements sont inclus dans A

- P(événement élémentaire) = 0
- La surface délimitée par la fonction de densité sur tout Ω est égale à 1
- Les valeurs prises par la fonction de densité sont ≥ 0 partout

Probabilités conditionnelles (1/5)

Définition

la probabilité d'un événement A conditionnellement à un événement B, que l'on note $P(A|B)$, est la probabilité que A se produise sachant que B s'est ou va se produire

Problème : comment calculer $P(A|B)$?

$P(A|\Omega) = P(A)$ puisqu'on sait que Ω sera réalisé

Probabilités conditionnelles (2/5)

Exemple

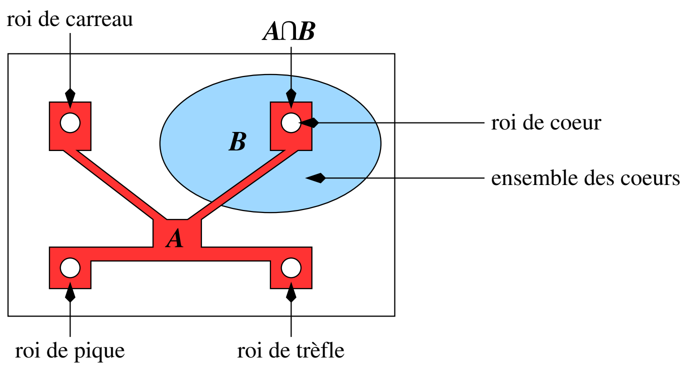
- Tirer une carte parmi un jeu de 32 cartes
- $\Omega = \{32 \text{ cartes}\}$
- événements : $A =$ tirer un roi $\qquad B =$ tirer un cœur
- $P(A|B)$ = ?

Si A se produit, sachant que B s'est aussi produit

$\Longrightarrow A \cap B$ se produit

$\Longrightarrow P(A|B) = P(A \cap B|B)$

Probabilités conditionnelles (3/5)



$\Longrightarrow$ on cherche la probabilité que le cercle $A \cap B$, qui est inclus dans B, soit réalisé, sachant que l'événement B est réalisé

$\Longrightarrow$ $P(B) = 1$ et B peut jouer le rôle d'univers pour $A \cap B$

Probabilités conditionnelles (4/5)

$$P(A|B) = \frac{|A \cap B|}{|B|}.$$

$$P(A|B) = \frac{|A \cap B|}{|\Omega|} \times \frac{|\Omega|}{|B|} = \frac{P(A \cap B)}{P(B)}.$$

Définition

$P(A|B) = \frac{P(A \cap B)}{P(B)}$ quand $P(B) > 0$

Probabilités conditionnelles (5/5)

$B_k,\ k \in K$ partition de Ω

$$A = A \cap \Omega = A \cap \left[\bigcup_{k \in K} B_k\right] = \bigcup_{k \in K} [A \cap B_k]$$

$$P(A) = \sum_{k \in K} P(A \cap B_k)$$

Or $P(A|B_k) = \frac{P(A \cap B_k)}{P(B_k)} \Longrightarrow P(A \cap B_k) = P(A|B_k)P(B_k)$

$$P(A) = \sum_{k \in K} P(A|B_k)P(B_k)$$

Application peu pratique

Paquet 1

Paquet 2

- Dans chaque paquet : des cartes gagnantes et des cartes perdantes
- Dans chaque paquet, tirer 1 carte marron $\Longrightarrow$ plus de chance d'avoir une carte gagnante que si on tire une carte bleue

Paquet 3 = paquet 1 $+$ paquet 2

Pb : Doit-on choisir une carte bleue ou marron dans le paquet 3 ?

Application peu pratique — suite

Paquet 1

	Bleue	Marron
Gagnante	1	2
Perdante	3	5

Paquet 2

	Bleue	Marron
Gagnante	5	3
Perdante	2	1

$$P(\text{Gagnante}|\text{Bleue}) = \begin{cases} 1/4 \text{ pour le paquet 1} \\ 5/7 \text{ pour le paquet 2} \end{cases}$$

$$P(\text{Gagnante}|\text{Marron}) = \begin{cases} 2/7 \text{ pour le paquet 1} \\ 3/4 \text{ pour le paquet 2} \end{cases}$$

Remarque : $1/4 < 2/7$ et $5/7 < 3/4$

$\Longrightarrow$ pour chaque paquet, choisir de préférence une carte marron

Application peu pratique — suite

Paquet 3

	Bleue	Marron
Gagnante	6	5
Perdante	5	6

$P(\text{Gagnante}|\text{Bleue}) = 6/11$

$P(\text{Gagnante}|\text{Marron}) = 5/11$

$\Longrightarrow$ pour ce paquet, choisir de préférence une carte bleue

Application peu pratique — fin

Paquet 1

	Bleue	Marron
G	1	2
P	3	5

Paquet 2

	Bleue	Marron
G	5	3
P	2	1

Paquet 3

	Bleue	Marron
G	6	5
P	5	6

Calculs pour le paquet 3 :

$$\begin{aligned}
P(\text{G}|\text{Bleue}) &= P(\text{G,Paquet 1}|\text{Bleue}) + P(\text{G,Paquet 2}|\text{Bleue}) \\
&= P(\text{G}|\text{Paquet 1,Bleue}) \times P(\text{Paquet 1}|\text{Bleue}) \\
&\quad + P(\text{G}|\text{Paquet 2,Bleue}) \times P(\text{Paquet 2}|\text{Bleue}) \\
&= P(\text{G}|\text{Paquet 1,Bleue}) \times 4/11 \\
&\quad + P(\text{G}|\text{Paquet 2,Bleue}) \times 7/11 \\
P(\text{G}|\text{Marron}) &= P(\text{G}|\text{Paquet 1,Marron}) \times 7/11 \\
&\quad + P(\text{G}|\text{Paquet 2,Marron}) \times 4/11
\end{aligned}$$

⚠ probas conditionnelles non pondérées par les mêmes poids

$\Longrightarrow$ choisir une carte bleue pour le paquet 3

Théorème de Bayes

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

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

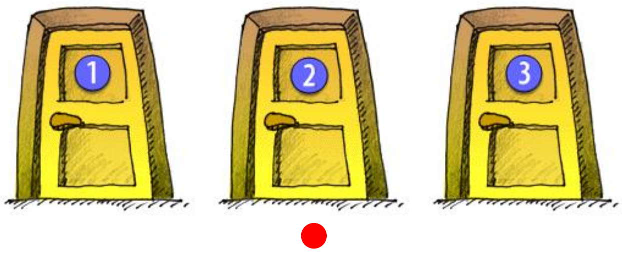
Théorème de Bayes

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad \forall A, B \text{ tels que } P(A) \neq 0 \text{ et } P(B) \neq 0$$

$$P(B|A) = \frac{P(A|B)P(B)}{\sum_B P(A|B)P(B)}$$

Monty Hall

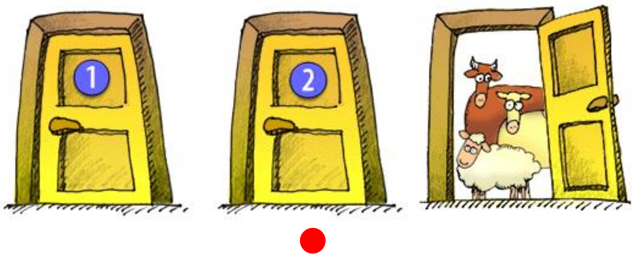
`http://www.apprendre-en-ligne.net/random/monty/`



- derrière une des portes = 1 bateau
- derrière les autres portes : des moutons

proba que le bateau se trouve derrière la porte ?

Monty Hall



L'animateur choisit au hasard une des 2 autres portes qui contient des moutons et l'ouvre.

Choix : conserver la porte 2 ou choisir la porte 1 ?

calculer la proba que le bateau soit derrière la porte 1 ou la 2

Monty Hall

3 événements élémentaires :

- B1 : le bateau est derrière la porte 1
- B2 : le bateau est derrière la porte 2
- B3 : le bateau est derrière la porte 3

E = événement « l'animateur a choisi, parmi les portes 1 et 3, d'ouvrir la porte 3 »

Choix $\Longrightarrow$ $P(B1|E)$ et $P(B2|E)$

Monty Hall

E = événement « l'animateur a choisi, parmi les portes 1 et 3, d'ouvrir la porte 3 »

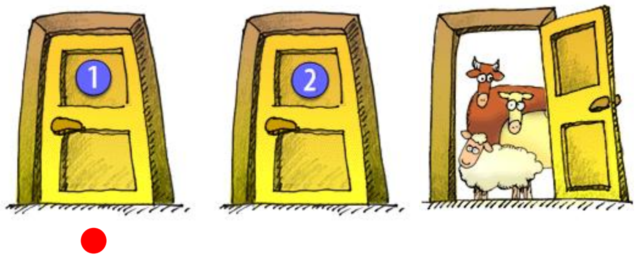
Formule de Bayes : $P(B1|E) = \dfrac{P(E|B1)P(B1)}{P(E)}$

- $P(E|B1) = 1$
- $P(B1) = 1/3$
- $P(E) = P(E|B1)P(B1) + P(E|B2)P(B2) + P(E|B3)P(B3)$
  $= 1 \times 1/3 + 1/2 \times 1/3 + 0 \times 1/3 = 1/2$

$$P(B1|E) = \frac{1 \times 1/3}{1/2} = 2/3$$

$P(B2|E) = 1 - P(B1|E) - P(B3|E) = 1/3$

Monty Hall



$P(B1|E) = 2/3$

$P(B2|E) = 1/3$

Indépendance probabiliste (1/2)

Définition de l'indépendance

deux événements A et B sont indépendants si :

$$P(A \cap B) = P(A) \times P(B)$$

si $P(B) > 0$ alors $P(A|B) = P(A)$

l'indépendance n'est pas une propriété du couple (A, B) mais du couple $(\{A, A^c\}, \{B, B^c\})$:

$$\begin{aligned} A \text{ et } B \text{ sont indépendants} &\Longrightarrow A \text{ et } B^c \text{ indépendants} \\ &\Longrightarrow A^c \text{ et } B \text{ indépendants} \\ &\Longrightarrow A^c \text{ et } B^c \text{ indépendants} \end{aligned}$$

Indépendance probabiliste (2/2)

Démonstration :

A et B sont indépendants $\Longrightarrow P(A \cap B) = P(A) \times P(B)$

$P(A) = P(A \cap B) + P(A \cap B^c)$

$$\begin{aligned} \Longrightarrow P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A) \times P(B) \\ &= P(A) \times [1 - P(B)] \\ &= P(A) \times P(B^c) \end{aligned}$$

Indépendance conditionnelle

$P(A|B) \approx$ restriction de l'espace Ω à B

Définition de l'indépendance conditionnelle

deux événements A et B sont indépendants conditionnellement à C si :

$$P(A \cap B|C) = P(A|C) \times P(B|C)$$

si $P(B \cap C) > 0$ alors $P(A|B, C) = P(A|C)$

l'indépendance conditionnelle n'est pas une propriété du couple (A, B) mais du couple $(\{A, A^c\}, \{B, B^c\})$:

A et B sont indépendants conditionnellement à C
$\Longrightarrow$ A et B^c indépendants conditionnellement à C
$\Longrightarrow$ A^c et B indépendants conditionnellement à C
$\Longrightarrow$ A^c et B^c indépendants conditionnellement à C

RFIDEC — cours 1 :
Rappels de probas/stats (3/3)

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours 1.3

1. variables aléatoires
2. caractéristiques et variables aléatoires
2. indépendances de variables aléatoires
3. loi binomiale et loi de Poisson
4. loi normale

Variables aléatoires (1/2)

Exemple

- au début : vous avez 1000 €
- jeu : je tire 3 cartes parmi un jeu de 32 cartes et, suivant le tirage, je vous donne ou vous prend de l'argent :

			
2000 €	500 €	0 €	-1000 €

$P(2000\,€) = P(3 \text{ rois})$

$P(\text{gain} \geq 500\,€) = P(3 \text{ rois}) + P(2 \text{ rois})$

Variables aléatoires (2/2)

Définition

- Ω = univers muni d'une loi de proba $P(\cdot)$
- Ω' un autre ensemble
- 2^{Ω} et $2^{\Omega'}$ = ensemble des sous-ensembles de Ω et de Ω'
- *variable aléatoire* = fonction Γ de 2^{Ω} dans $2^{\Omega'}$ telle que :

$$\Gamma^{-1}(A') \in 2^{\Omega} \quad \forall A' \in 2^{\Omega'}.$$

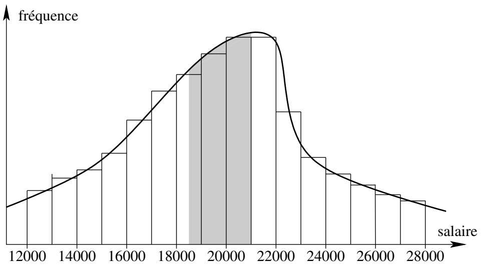
Probabilité sur Ω'

loi de proba $P'(\cdot)$ sur Ω' :

$$P'(A') = P\left(\Gamma^{-1}(A')\right) \quad \forall A' \in 2^{\Omega'}.$$

notation : $P(\Gamma = A') = P(\Gamma^{-1}(A'))$

Probabilités : retour sur le cas continu (1/2)



$$P(X \in I) = \int_I p(x)dx$$

avec P = proba et p = fonction de densité

$\Longrightarrow$ connaître p = connaître P

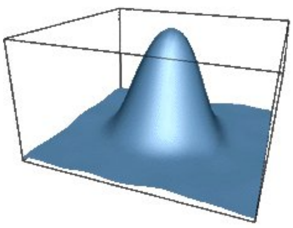
intervalles $]-\infty, x[$ $\Longrightarrow$ fonction de répartition :

$$F(x) = P(X < x) = \int_{-\infty}^{x} p(y)dy$$

Probabilités : retour sur le cas continu (2/2)

- variables bi-dimensionnelles, ou couples de variables, continues, (X, Y)
- densité de probabilité = $p(x, y)$
- fonction de répartition = $F(x, y)$
- alors $\forall x, y$:

$$F(x,y) = P(X < x, Y < y) = \int\int_{\{(x',y'):x'<x,y'<y\}} p(x',y')dx'dy'$$



Caractéristiques d'une loi de probabilités (1/3)

Caractéristiques

- Médiane M : $P(X \leq M) \geq \frac{1}{2}$ et $P(X \geq M) \geq \frac{1}{2}$
- Espérance mathématique ou moyenne : $E(X)$

  X discrète : $E(X) = \sum x_k p_k$

  X continue : $E(X) = \displaystyle\int x\, p(x) dx$

⚠ l'espérance mathématique n'existe pas toujours

- Mode : Mo de P (pas toujours unique) :

  X discrète : $p(Mo) = \max_k p(x_k)$

  X continue : $p(Mo) = \max_x p(x)$

Propriétés de l'espérance

- $E(aX + b) = aE(X) + b$
- $\forall X, Y,\ E(X + Y) = E(X) + E(Y)$

Application de l'espérance (1/3)

- X = variable aléatoire $\approx$ 1 caractère dans une phrase
- domaine de $X = \{x_1, \ldots, x_{16}\}$

Encodage « classique d'une phrase » :

- 16 valeurs possibles $\Longrightarrow$ codage sur 4 bits

  $\Longrightarrow$ phrase de 10 caractères = 40 bits

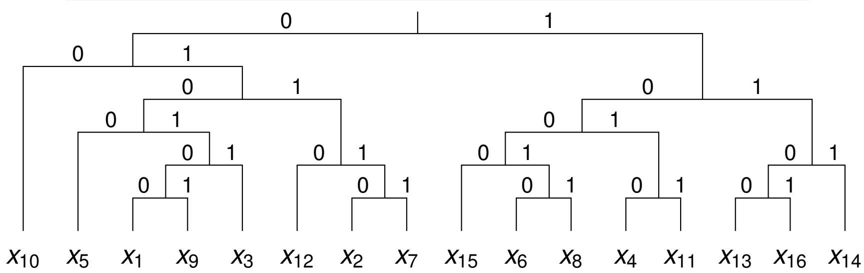
Proba d'apparition des caractères ($\times$100)															
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
1	3	2	6	5	4	2	3	1	21	6	5	9	17	7	8

- Minimisation de l'espérance du nombre de bits $\Longrightarrow$ compression

Application de l'espérance (2/3)

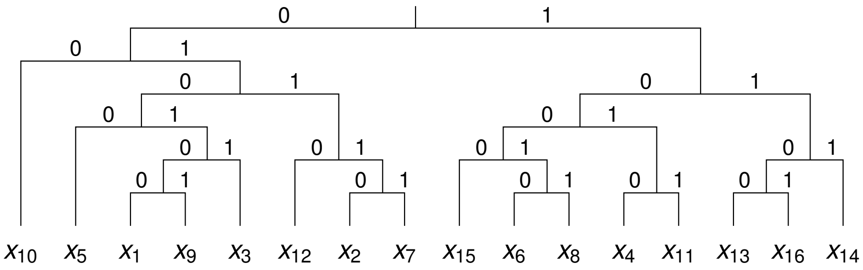
Proba d'apparition des caractères ($\times$100)															
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
1	3	2	6	5	4	2	3	1	21	6	5	9	17	7	8

Codage de Huffman



$\Longrightarrow$ par exemple $x_3 = 01011$

Application de l'espérance (3/3)



Nombre de bits et proba ($\times 100$) par caractère																
X_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
nb bits	6	5	5	4	4	5	5	5	6	2	4	4	4	3	4	4
proba	1	3	2	6	5	4	2	3	1	21	6	5	9	17	7	8

$$\text{Espérance du nombre de bits} = \sum_{i=1}^{16} p_i |\text{nb bits } x_i| = 3,59 < 4$$

Caractéristiques d'une loi de probabilités (2/3)

Caractéristiques de dispersion

- variance : $V(X)$ ou σ^2 :

  X discrète : $\sigma^2 = \sum [x_k - E(X)]^2 p_k$

  X continue : $\sigma^2 = \int [x - E(X)]^2\, p(x)dx$
- moyenne des carrés des écarts entre les valeurs prises par X et son espérance $E(X)$
- écart-type : σ = racine carrée de la variance
- variance et donc écart-type n'existent pas toujours

Caractéristiques d'une loi de probabilités (3/3)

Propriétés de la variance

- $V(X) = E(X^2) - E(X)^2$
- $\forall X, Y,\ V(X + Y) = V(X) + V(Y) + 2cov(X, Y)$ où :
  - $cov(X, Y)$ = covariance de X et Y
  - si X et Y discrètes et $p_k = P(X = x_k, Y = y_k)$

    $$cov(X, Y) = \sum [x_k - E(X)][y_k - E(Y)]p_k$$
  - si X et Y continues, de densité $p(x, y)$,

    $$cov(X, Y) = \iint [x - E(X)][y - E(Y)]p(x, y)dxdy$$

Indépendance de deux variables aléatoires (1/3)

Indépendance de deux variables discrètes

X et Y sont *indépendantes* si $\forall x, \forall y$:

les événements $X = x$ et $Y = y$ sont indépendants

- $\forall x, \forall y,\ P(X = x \cap Y = y) = P(X = x) \times P(Y = y)$
- $\forall x, \forall y\ t.q.\ P(Y = y) > 0,\ P(X = x|Y = y) = P(X = x)$
- $\forall y, \forall x\ t.q.\ P(X = x) > 0,\ P(Y = y|X = x) = P(Y = y)$

Indépendance de deux variables aléatoires (2/3)

Indépendance de deux variables continues

X et Y sont *indépendantes* si $\forall I, \forall J$, intervalles,

les événements $X \in I$ et $Y \in J$ sont indépendants

Il suffit que les fonctions de répartition, F_X, F_Y de X et Y et F_{XY} du couple satisfassent :

$$\forall x, y,\ F_{XY}(x, y) = F_X(x) \times F_Y(y)$$

ou encore que les densités de probabilité p_X, p_Y de X et Y et p_{XY} du couple satisfassent :

$$\forall x, y,\ p_{XY}(x, y) = p_X(x) \times p_Y(y)$$

Indépendance de deux variables aléatoires (3/3)

Propriété

La covariance de deux variables indépendantes X et Y est toujours nulle

La réciproque est fausse

$$cov(X, Y) = \iint [x - E(X)][y - E(Y)]p_{XY}(x, y)dxdy$$

X et Y indep $\Longrightarrow p_{XY}(x, y) = p_X(x) \times p_Y(y)$

$$\Longrightarrow cov(X, Y) = \int [x - E(X)]p_X(x)dx \int [y - E(Y)]p_Y(y)dy$$

Or $\int [x - E(X)]p_X(x)dx = \int xp_X(x)dx - \int E(X)p_X(x)dx$

$= E(X) - E(X) = 0$

$\Longrightarrow cov(X, Y) = 0 \times 0 = 0$

Indépendance mutuelle de n variables (1/3)

Définition

n variables $(X_1, X_2, \ldots, X_k, \ldots, X_n)$

Elles sont *mutuellement indépendantes* si tout événement lié à une partie d'entre elles est indépendant de tout événement lié à toute autre partie disjointe de la précédente

$\Longrightarrow$ c'est la généralisation naturelle de l'indépendance de deux variables :

des variables discrètes $(X_1, \ldots, X_k, \ldots, X_n)$ sont mutuellement indépendantes lorsque :

$$\forall x_k \;\; P\left(\bigcap_{k=1}^{n} X_k = x_k\right) = \prod_{k=1}^{n} P(X_k = x_k)$$

Indépendance mutuelle de n variables (2/3)

des variables continues $(X_1, \ldots, X_k, \ldots, X_n)$ sont mutuellement indépendantes lorsque $\forall x_k \; (k = 1, \ldots, n)$:

$$p_{X_1 \ldots X_k \ldots X_n}(x_1, \ldots, x_k, \ldots, x_n) = \prod_{k=1}^{n} p_{X_k}(x_k)$$

L'indépendance mutuelle de n variables entraîne leur indépendance deux à deux

la réciproque n'est pas vraie

Indépendance mutuelle de n variables (3/3)

- Soit n dés à 6 faces
- X_k : variable aléatoire indiquant sur quelle face tombe le $k$ème dé
- tous les dés sont différents $\Longrightarrow X_1, \ldots, X_n$ mutuellement indépendantes

$$p_{X_1 \ldots X_k \ldots X_n}(x_1, \ldots, x_k, \ldots, x_n) = \prod_{k=1}^{n} p_{X_k}(x_k)$$

$\Longrightarrow$ stockage mémoire = $6n$ au lieu de 6^n

n = 10	60	60 millions
n = 20	120	3,6 millions de milliards

Indépendance conditionnelle (1/2)

Indépendance conditionnelle de deux variables discrètes

X et Y sont *indépendantes conditionnellement à* Z si $\forall x, \forall y, \forall z$, les événements $X = x$ et $Y = y$ sont indépendants conditionnellement à $Z = z$

- $P(X=x \cap Y=y|Z=z) = P(X=x|Z=z) \times P(Y=y|Z=z)$
- si $P(Y=y|Z=z) > 0$ alors :
  $P(X=x|Y=y, Z=z) = P(X=x|Z=z)$
- si $P(X=x|Z=z) > 0$ alors :
  $P(Y=y|X=x, Z=z) = P(Y=y|Z=z)$

Indépendance conditionnelle (2/2)

Indépendance conditionnelle de deux variables discrètes

X et Y sont *indépendantes* conditionnellement à Z si :

- $P(X \cap Y|Z) = P(X|Z) \times P(Y|Z)$
- si $P(Y|Z) > 0$ alors $P(X|Y, Z) = P(X|Z)$
- si $P(X|Z) > 0$ alors $P(Y|X, Z) = P(Y|Z)$

Interprétation

- Conditionnement = apport de connaissances
- Si l'on connaît la valeur de la variable Z, alors connaître celle de Y n'apporte rien sur la connaissance de X

⚠ Ces formules s'étendent si X, Y et/ou Z sont remplacés par des ensembles de variables aléatoires disjoints 2 à 2

Application des probas conditionnelles (1/2)

- Classe de 40 étudiants
- assertion : ≪il y a au moins 2 étudiants qui sont nés le même jour≫

Avez-vous intérêt à parier 10 € que cette assertion est vraie ?

- $X_i \in \{1, \ldots, 365\}$ le jour de naissance du $i$ème étudiant
- que vaut $P(\text{tous les } X_i \text{ sont différents})$?

Application des probas conditionnelles (2/2)

$$\alpha = P(\text{tous les } X_i \text{ sont différents})$$
$$= P(X_2 \neq X_1) \times P(X_3 \notin \{X_1, X_2\} | X_2 \neq X_1) \times$$
$$P(X_4 \notin \{X_1, X_2, X_3\} | X_2 \neq X_1, X_3 \notin \{X_1, X_2\}) \times \dots$$
$$= \prod_{i=2}^{40} P\left(X_i \notin \{X_j : j < i\} \middle| \bigwedge_{j<i} X_j \notin \{X_k : k < j\} \right)$$
$$= \prod_{i=2}^{40} \frac{365 - (i-1)}{365} = \prod_{i=1}^{39} \frac{365 - i}{365} \approx 10,87\%$$

$\Longrightarrow$ en choisissant au hasard une classe de 40 étudiants,

on a 10,87% de chances que l'assertion soit fausse

Loi de Bernoulli

Définition

Épreuve de Bernoulli = expérience aléatoire qui ne peut prendre que deux résultats (*succès* et *échec*)

p = proba de succès, et $q = 1 - p$ = proba d'échec.

Loi de Bernoulli

Variable X à support $\mathcal{X} = \{0, 1\}$ telle que :

$P(X = 1) = p$ et $P(X = 0) = 1 - p$

$E(X) = p \quad V(X) = p(1-p)$

$\Longrightarrow X =$ le nombre de succès de l'épreuve de Bernoulli

Loi binomiale

Définition

Épreuve binomiale = expérience aléatoire telle que :

1. on répète n fois la même épreuve de Bernoulli,
2. les probas p et q restent inchangées pour chaque épreuve de Bernoulli,
3. les épreuves de Bernoulli sont toutes réalisées indépendamment les unes des autres.

Loi binomiale de paramètres n et p

- X = nombre de succès de l'épreuve binomiale
- $X \sim \mathcal{B}(n, p)$
- $P(X = k) = C_n^k \, p^k \, (1-p)^{n-k}, \forall k = 0, \dots, n$
- $E(X) = np \quad V(X) = np(1-p)$

Loi de Poisson

Définition

- *Loi de Poisson* ou loi des événements rares :
  loi vers laquelle tend la loi binomiale lorsque n est très grand et p assez petit
- variable discrète $X \sim$ *loi de* POISSON de paramètre $\lambda > 0$:
  $\Longrightarrow$ les événements élémentaires $e_k = \ll X = k \gg$, $k \in \mathbb{N}$, ont pour probabilités :

$$P(X = k) = \frac{e^{-\lambda} \times \lambda^k}{k!}$$

- $E(X) = \lambda \quad V(X) = \lambda$

$\Longrightarrow$ s'applique dans le cas d'expériences dont les occurrences sont totalement aléatoires (sans régularité) et indépendantes les unes des autres

Lois géométrique et exponentielle

Définition : loi géométrique de paramètre p

- loi de la variable donnant le nombre d'essais nécessaires pour que se produise un événement de probabilité p
- $P(X = k) = p(1-p)^{(k-1)}$, $k = 1, \dots, n, \dots$
- $E(X) = \frac{1}{p} \quad V(X) = \frac{1-p}{p^2}$

Définition : loi exponentielle de paramètre λ

- a pour support $\mathbb{N}^*_+$ et pour densité
  $p(x) = \lambda \exp\{-\lambda x\}$, $x > 0$
- $E(X) = \frac{1}{\lambda} \quad V(X) = \frac{1}{\lambda^2}$
- utilisée en fiabilité : durée de vie de nombreux composants
  λ = *taux de défaillance*
  $E(X) = \frac{1}{\lambda}$ = *temps moyen entre défaillances*

Loi normale

⚠ Loi extrêmement importante : souvent une très bonne approximation de la loi réelle

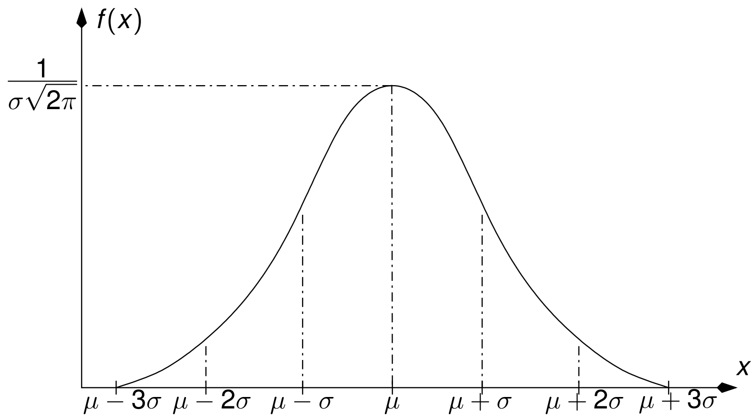
Définition : loi normale de paramètres μ et σ^2

- notée $\mathcal{N}(\mu, \sigma^2)$
- s'applique pour des variables aléatoires continues
- densité positive sur tout $\mathbb{R}$:

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

- $E(X) = \mu \quad V(X) = \sigma^2$

Fonction de densité de la loi normale

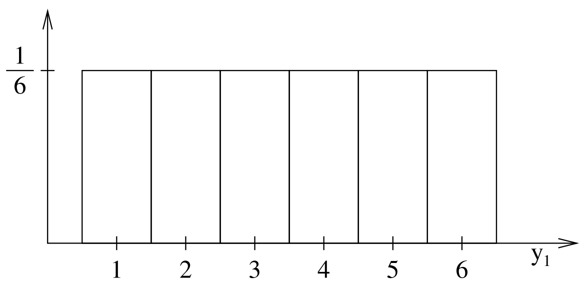


Loi normale = limite d'autres lois (1/4)

Lancés de dés à 6 faces

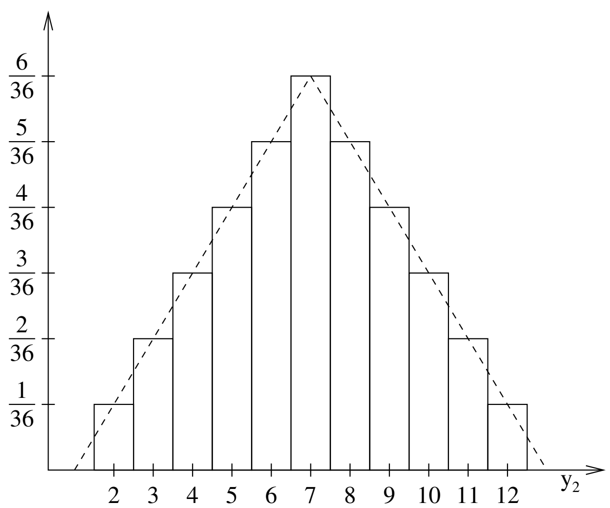
$\Longrightarrow$ on compte la somme des résultats des dés

Somme pour 1 jet de dé



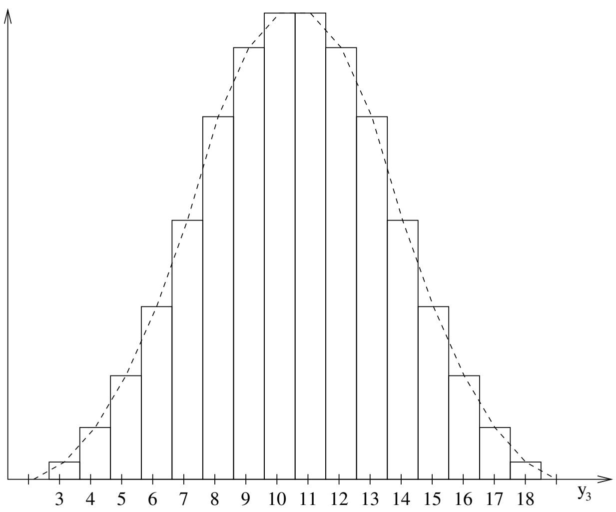
Loi normale = limite d'autres lois (2/4)

Somme pour 2 jets de dés



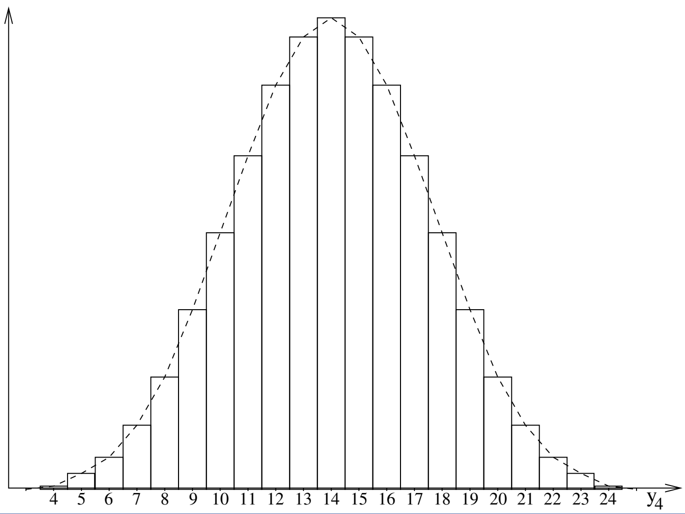
Loi normale = limite d'autres lois (3/4)

Somme pour 3 jets de dés



Loi normale = limite d'autres lois (4/4)

Somme pour 4 jets de dés



Loi normale en pratique

Théorème

$X \sim \mathcal{N}(\mu; \sigma^2)$

Alors la variable $Y = aX + b$ obéit à la loi $\mathcal{N}(a\mu + b; a^2\sigma^2)$.

$\Longrightarrow$ toute transformée affine d'une variable aléatoire suivant une loi normale suit aussi une loi normale

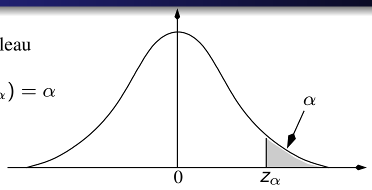
Corollaire

- X une variable aléatoire obéissant à une loi $\mathcal{N}(\mu; \sigma^2)$

  $\Longrightarrow Z = \dfrac{X - \mu}{\sigma}$ suit la loi $\mathcal{N}(0; 1)$
- Z suit une loi normale centrée (à cause de la moyenne en 0) réduite (à cause du σ^2 égal à 1)

Table de la loi normale centrée réduite

valeurs dans le tableau
ci-dessous : les α
tels que $P(Z > z_\alpha) = \alpha$



z_α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0859	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0722	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0466	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

Loi normale bi-dimensionnelle

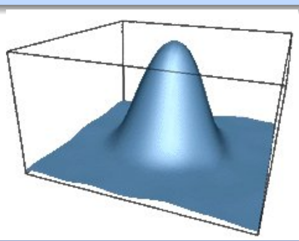
Définition : loi normale bi-dimensionnelle

- couple de variables (X, Y)
- densité dans $\mathbb{R}^2$:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

où $\rho = \frac{cov(X,Y)}{\sigma_x\sigma_y}$ = coefficient de corrélation linéaire



RFIDEC — cours 2 : Échantillons, estimations ponctuelles

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours n˚2

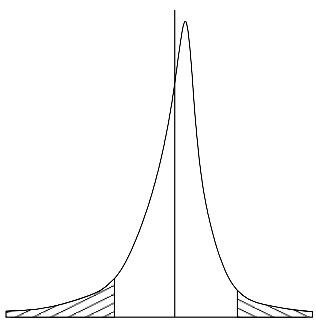
1. Lois des grands nombres
2. Théorème central-limite
3. Estimation ponctuelle à partir d'échantillons
4. Biais dans les estimations

Convergence en probabilité

Définition

- $(X_n)_{n \in \mathbb{N}}$: suite de variables
- a : constante
- (X_n) *converge en probabilité* vers a si, pour tout $\epsilon > 0$ la probabilité que l'écart absolu entre X_n et a dépasse ϵ tend vers 0 quand $n \to \infty$:

$$\lim_{n \to \infty} P(|X_n - a| \geq \epsilon) = 0$$



Aire hachurée tend vers 0 quand $n \to \infty$

Loi faible des grands nombres

Loi faible

- $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires :
  - de même loi
  - d'espérance m
  - possédant une variance σ^2
  - deux à deux indépendantes
- alors la suite des variables $\overline{X}_n = \frac{\sum_{k=1}^{n} X_k}{n}$ converge en probabilité vers m

$\overline{X}_n$ est appelée *moyenne empirique*

$E(\overline{X}_n) = m$

$V(\overline{X}_n) = \frac{\sigma^2}{n}$

conséquence : échantillons de grandes tailles $\Longrightarrow$ bonne chance d'estimer m

Convergence presque sûre

Définition

- $(X_n)_{n\in\mathbb{N}}$: suite de variables
- a : constante
- (X_n) *converge presque sûrement* vers a s'il y a une proba 1 que la suite des réalisations des X_n tende vers a :

$$P\left(\lim_{n\to\infty} X_n = a\right) = 1$$

⚠ Définition plus exigeante que la convergence en probabilité

Loi forte des grands nombres

Loi forte

- $(X_n)_{n\in\mathbb{N}}$: suite de variables aléatoires
  - de même loi
  - d'espérance m
  - possédant une variance σ^2
  - mutuellement indépendantes
- alors la suite des variables $\overline{X}_n = \frac{\sum_{k=1}^{n} X_k}{n}$ converge presque sûrement vers m

Interprétation : échantillon de grande taille $\Longrightarrow$ bonne estimation de m

Convergence en loi

Définition

- $(X_n)_{n\in\mathbb{N}}$: suite de variables
- F_n : fonction de répartition de X_n
- X : variable de fonction de répartition F
- La suite X_n *converge en loi* vers X lorsque $F_n(x)$ tend vers $F(x)$ en tout point de continuité de F

Notation : $X_n \overset{loi}{\rightarrow} X$

Théorème central-limite

Théorème central-limite

- $(X_n)_{n\in\mathbb{N}}$: suite de variables
  - de même loi
  - d'espérance μ
  - de variance σ^2
  - mutuellement indépendantes
- alors la suite des moyennes empiriques centrées réduites $\frac{\overline{X}_n-\mu}{\sigma\,/\,\sqrt{n}}$ tend en loi vers la loi normale centrée réduite :

$$\frac{\overline{X}_n-\mu}{\sigma\,/\,\sqrt{n}} \overset{loi}{\rightarrow} \mathcal{N}(0,1)$$

Illustration du théorème central-limite (1/4)

Lancés de dés à 6 faces 

$$\Longrightarrow \begin{cases} X_i & = \text{résultat du jet du } i\text{ème dé} \\ \overline{X}_n & = \text{somme des résultats des dés} \end{cases}$$

distribution de $\overline{X}_n$ pour 1 jet de dé

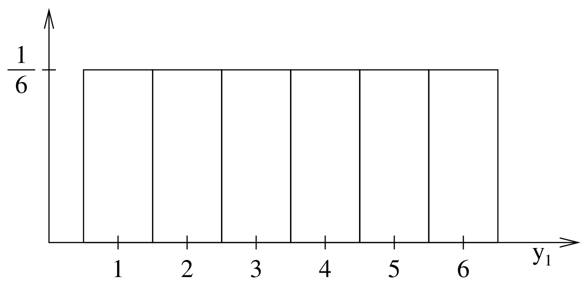


Illustration du théorème central-limite (2/4)

distribution de $\overline{X}_n$ pour 2 jets de dés

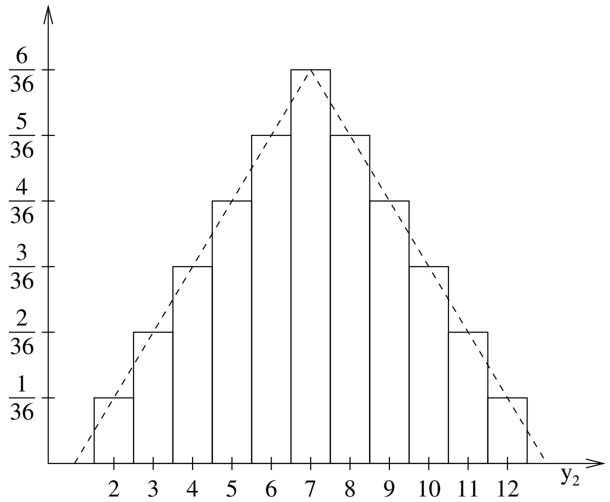


Illustration du théorème central-limite (3/4)

distribution de $\overline{X}_n$ pour 3 jets de dés

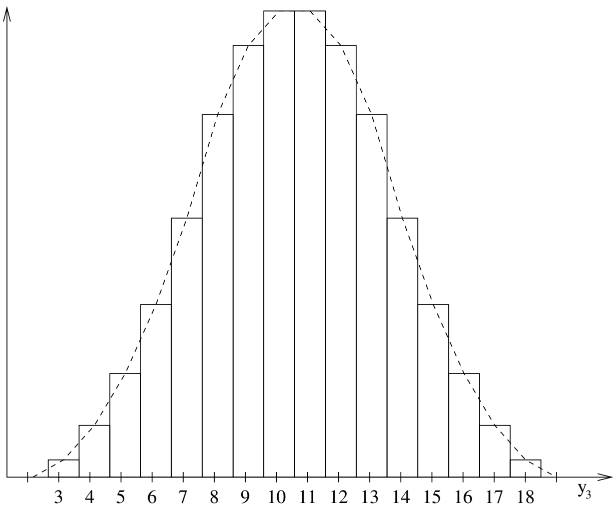
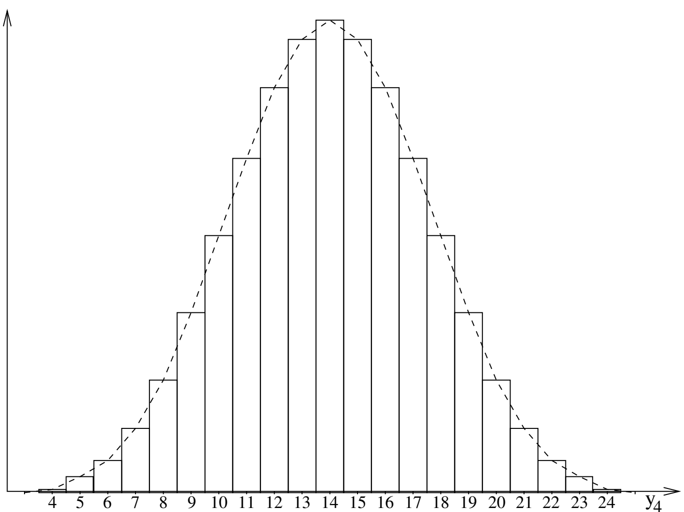


Illustration du théorème central-limite (4/4)

distribution de $\overline{X}_n$ pour 4 jets de dés



Statistique inférentielle

Statistique inférentielle

Hypothèses :

- les données = les observations
- observations $(x_1, \dots, x_n)$ = réalisation d'une variable aléatoire multidimensionnelle $X = (X_1, \dots, X_n)$
- observations = *échantillon empirique*
- l'échantillon est tiré suivant une loi X_0
- chaque variable X_i suit la même loi que X_0

But :

- déduire des informations sur X_0

Statistique inférentielle

Exemples :

- sondages électoraux, études de marché
- tests de fiabilité/qualité
- bases de données $\Longrightarrow$ diagnostic

Estimation ponctuelle

Idée force

Dans moult études statistiques :

- population de grande taille
  $\Longrightarrow$ impossible de la connaître précisément
- possibilité de prélever des échantillons

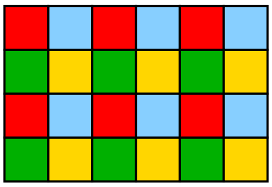
$\Longrightarrow$ étudier l'échantillon et en déduire les caractéristiques de la population

caractéristiques : moyenne μ, variance σ^2, proportion de succès p

Échantillonnage (1/2)

caractéristique de la population $\Longleftarrow$ échantillon

$\Longrightarrow$ problème : comment prélever l'échantillon ?



Définition d'un échantillon

Échantillon d'une population = sous-ensemble représentatif de la population
$\Longrightarrow$ composition de l'échantillon due au hasard

Échantillonnage (2/2)

Prélèvement d'un échantillon

2 manières de prélever un échantillon aléatoirement :

1. Échantillonnage avec remise :
   choisir un individu au hasard
   noter la valeur de la variable d'intérêt pour celui-ci
   remettre l'individu dans la population
   réitérer le processus
2. Échantillonnage sans remise :
   même procédé mais sans remettre dans la population les individus sélectionnés

⚠ échantillonnage avec remise $\Longrightarrow$ un individu peut apparaître plusieurs fois dans l'échantillon

⚠ population de grande taille $\Longrightarrow$ avec remise $\approx$ sans remise

Estimation de la moyenne d'une population (1/4)

⚠ résultats suivant valables uniquement pour des échantillons *avec remise*

$\Longrightarrow$ on va travailler avec des échantillons i.i.d :

Échantillon i.i.d

- échantillon de n individus
- X_i = variable aléatoire « valeur du $i$ème individu tiré »
- les X_i sont mutuellement indépendants
- les X_i ont tous la même distribution

$\Longrightarrow$ les X_i sont indépendants et identiquement distribués (i.i.d)

échantillons i.i.d $\Longrightarrow$ bonnes propriétés mathématiques

Estimation de la moyenne d'une population (2/4)

- X variable aléatoire sur l'ensemble de la population
- espérance : μ, variance : σ^2

- échantillon de n individus $\Longrightarrow$ observation de n valeurs de X
- X_i : variable aléatoire correspondant au $i$ème individu
- échantillon i.i.d $\Longrightarrow$ espérance de X_i : μ, variance de X_i : σ^2

- $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ = moyenne de l'échantillon

Problèmes : que valent $E(\overline{X})$ et $V(\overline{X})$?

Estimation de la moyenne d'une population (3/4)

- $$\begin{aligned} E(\overline{X}) &= \frac{1}{n} E(X_1 + X_2 + \cdots + X_n) \\ &= \frac{1}{n}\left(E(X_1) + E(X_2) + \cdots + E(X_n)\right) \\ &= \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu \end{aligned}$$

variables X_i mutuellement indépendantes $\Longrightarrow$

- $$\begin{aligned} V(\overline{X}) = V\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) &= \frac{1}{n^2} V(X_1 + X_2 + \cdots + X_n) \\ &= \frac{1}{n^2}\left(V(X_1) + V(X_2) + \cdots + V(X_n)\right) \\ &= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{\sigma^2}{n} \end{aligned}$$

Estimation de la moyenne d'une population (4/4)

Théorème 1

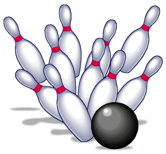
- X : variable aléatoire
- espérance de X : μ, variance de X : σ^2
- échantillon de taille n avec remise sur X
- $\overline{X}$: variable aléatoire « moyenne de l'échantillon »
- Alors : $E(\overline{X}) = \mu$ et $V(\overline{X}) = \frac{\sigma^2}{n}$

Corollaire

- X : variable aléatoire suivant une loi normale $\mathcal{N}(\mu; \sigma^2)$
- échantillon de taille n avec remise sur X
- $\overline{X}$: variable aléatoire « moyenne de l'échantillon »
- Alors : $\overline{X} \sim \mathcal{N}(\mu; \frac{\sigma^2}{n})$

Exemple : calcul du handicap au bowling

Compétition en 2 phases :



- 1ère phase : 6 parties
- calcul du handicap :
  Handicap H = (215 $-$ moyenne des 6 parties)$\times$60%
- 2ème phase : 6 parties (score $+$ le handicap H)

- Calcul du handicap $\Longrightarrow$ estimation de votre niveau
- 6 parties de la 1ère phase = échantillon de taille 6
- X_i = variable aléatoire « score de la $i$ème partie »

idée : scores de l'échantillon $\Longrightarrow$ score moyen de la population

Retour sur l'estimation de la moyenne

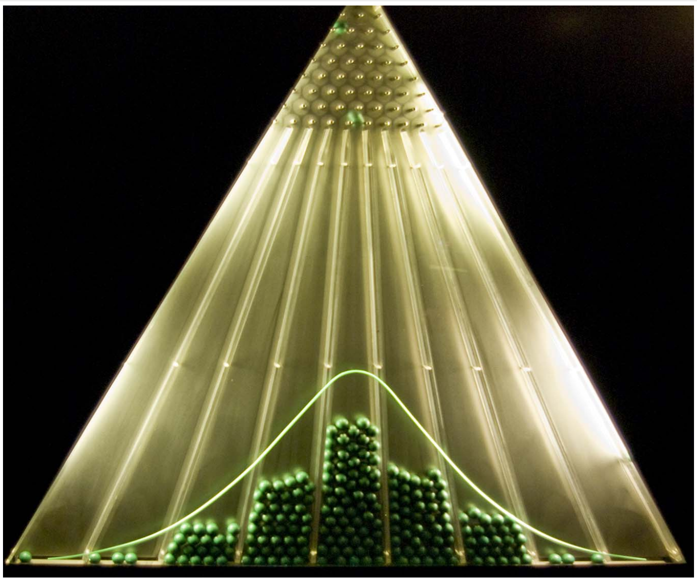
Problème : que faire si X ne suit pas une loi normale ?

$\Longrightarrow$ sauvé grâce au théorème central-limite :

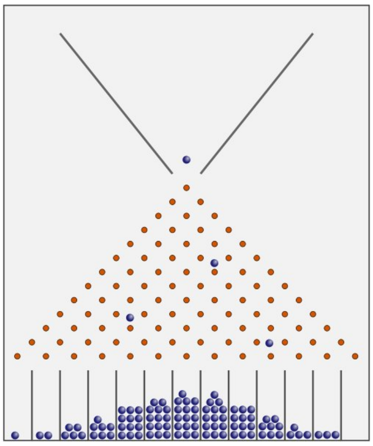
Théorème 2

- X : variable aléatoire
- espérance de X : μ, variance de X : σ^2
- échantillon de taille n avec remise sur X
- n suffisamment grand ($n \geq 30$ si la distribution de X n'est pas trop dissymétrique, $n \geq 50$ sinon)
- $\overline{X}$: variable aléatoire « moyenne de l'échantillon »
- Alors : $\dfrac{\overline{X} - E(\overline{X})}{\sqrt{V(\overline{X})}} = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0; 1)$

Exemple 1 : la planche de Galton (1/2)



Exemple 1 : la planche de Galton (2/2)



- chaque niveau $\Longrightarrow$ expérience de Bernoulli
- $\Longrightarrow X \sim$ loi binomiale

  $\Longrightarrow X \not\sim$ loi normale
- théorème 2 $\Longrightarrow$

$$\frac{\overline{X} - E(\overline{X})}{\sqrt{V(\overline{X})}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0; 1)$$

Exemple 2 : analyse des déchets

- Grenelle de l'environnement

  $\Longrightarrow$ réduction des déchets

  $\Longrightarrow$ analyse des déchets
- impossible à réaliser sur toute la population

  $\Longrightarrow$ échantillon de taille 100 :

450	320	320	390	410	415	380	390
440	350	400	380	430	400	375	...

- moyenne de l'échantillon $=$ 390 kg/an/habitant
- $\overline{X}$: variable aléatoire « moyenne de l'échantillon »
- $\sigma = 20$ supposé connu
- $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \Longrightarrow$ écart-type de $\overline{X} = 20/10 = 2$

Estimation de proportions de succès (1/3)

Problème :  v.s. 

- lancement de la Wii $\Longrightarrow$ étude de marché
- échantillon de taille $n \Longrightarrow$ succès = Wii, échec = PS3
- $p =$ proportion de succès dans toute la population
- P_i : variable aléatoire « succès du $i$ème individu »

Question : peut-on déduire p en observant les p_i ?

Estimation de proportions de succès (2/3)

Question : peut-on déduire p en observant les p_i ?

- p = proportion de succès dans la population (p pas trop petit)
- échantillon i.i.d de taille n assez grand
- P_i = variable aléatoire « succès du $i$ème individu »
- $P_i \sim$ loi binomiale $\mathcal{B}(1, p)$
- $\overline{P}$ = moyenne de l'échantillon

Théorème central-limite $\Longrightarrow \dfrac{\overline{P} - E(\overline{P})}{\sqrt{V(\overline{P})}} \sim \mathcal{N}(0; 1)$

Or $\overline{P} \sim \frac{1}{n}\mathcal{B}(n; p) \Longrightarrow E(\overline{P}) = \dfrac{np}{n}$ et $V(\overline{P}) = \dfrac{p(1-p)}{n}$

Estimation de proportions de succès (3/3)

Théorème 3

- échantillon i.i.d de taille n assez grand
- p = proportion de succès dans la population (p pas trop petit)
- $\overline{P}$ = moyenne de l'échantillon
- Alors : $\dfrac{\overline{P} - E(\overline{P})}{\sqrt{V(\overline{P})}} = \dfrac{\overline{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0; 1)$

$\Longrightarrow$ on peut estimer p en observant la valeur de $\overline{P}$

Application du théorème précédent

Expert : 80%  v.s. 20% 

- échantillon de 100 personnes $\Longrightarrow$ 70 Wii et 30 PS3
- Doit-on croire l'expert ?

si $p = 80\%$ alors Théorème 3 $\Longrightarrow \dfrac{\overline{P} - 0,8}{\sqrt{\frac{0,8 \times 0,2}{100}}} = 25(\overline{P} - 0,8) \sim \mathcal{N}(0; 1)$

$$\begin{aligned} \text{Prob}(\overline{P} \leq 0.7) &= \text{Prob}(25(\overline{P} - 0.8) \leq 25 \times (0.7 - 0.8)) \\ &= \text{Prob}(25(\overline{P} - 0.8) \leq -2,5) \approx 0,62\% \end{aligned}$$

Application au réchauffement climatique

- Avril 2007 : étude tms-sofres / CNRS : opinion des gens sur le réchauffement climatique
- 1000 personnes de 15 ans et + interrogées $\Longrightarrow$ échantillon i.i.d

- 790 pensent qu'il y a un changement climatique
- 210 ne le pensent pas
- $\overline{P}$: proportion de succès moyenne de l'échantillon
- p : proportion de personnes pensant qu'il y a dérèglement climatique dans la population française

$$\frac{\overline{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0;1)$$

$\Longrightarrow$ estimation de $p = 79\%$, écart-type de $\overline{P} \leq \sqrt{\frac{0.25}{1000}} \approx 0,015$

Caractéristiques des estimateurs

Transparents précédents $\Longrightarrow$ si échantillon de grande taille alors $\overline{X} \approx \mu \Longrightarrow$ estimer μ par $\overline{X}$

Estimateur non biaisé

- Estimateur T d'un paramètre $\theta \Longrightarrow$ valeur estimée $\hat{\theta}$
- T non biaisé si $E(T) = \theta$

⚠ estimateur biaisé $\Longrightarrow$ on surévalue ou sous-évalue θ

Estimateur convergent

- Estimateur T d'un paramètre θ
- T convergent si $E[(T-\theta)^2] \to 0$ lorsque la taille de l'échantillon $\nearrow$

moyenne $\overline{X}$ et proportion de succès $\overline{P}$: estimateurs non biaisés et convergents

Biais : estimation ponctuelle d'une variance (1/5)

- population $= 4$ nombres $\{1,2,3,4\}$
- $\mu = \frac{5}{2}$ et $\sigma^2 = \frac{5}{4}$
- échantillon de taille 2 avec remise :

échant.	espérance	variance	échant.	espérance	variance
1 1	$E(\overline{X}) = 1$	$V(\overline{X}) = 0$	1 2	$E(\overline{X}) = \frac{3}{2}$	$V(\overline{X}) = \frac{1}{4}$
1 3	$E(\overline{X}) = 2$	$V(\overline{X}) = 1$	1 4	$E(\overline{X}) = \frac{5}{2}$	$V(\overline{X}) = \frac{9}{4}$
2 1	$E(\overline{X}) = \frac{3}{2}$	$V(\overline{X}) = \frac{1}{4}$	2 2	$E(\overline{X}) = 2$	$V(\overline{X}) = 0$
2 3	$E(\overline{X}) = \frac{5}{2}$	$V(\overline{X}) = \frac{1}{4}$	2 4	$E(\overline{X}) = 3$	$V(\overline{X}) = 1$
3 1	$E(\overline{X}) = 2$	$V(\overline{X}) = 1$	3 2	$E(\overline{X}) = \frac{5}{2}$	$V(\overline{X}) = \frac{1}{4}$
3 3	$E(\overline{X}) = 3$	$V(\overline{X}) = 0$	3 4	$E(\overline{X}) = \frac{7}{2}$	$V(\overline{X}) = \frac{1}{4}$
4 1	$E(\overline{X}) = \frac{5}{2}$	$V(\overline{X}) = \frac{9}{4}$	4 2	$E(\overline{X}) = 3$	$V(\overline{X}) = 1$
4 3	$E(\overline{X}) = \frac{7}{2}$	$V(\overline{X}) = \frac{1}{4}$	4 4	$E(\overline{X}) = 4$	$V(\overline{X}) = 0$

- espérance $\begin{cases} \text{des moyennes des échantillons} = \mu \\ \text{des variances des échantillons} = \frac{5}{8} \neq \frac{5}{4} \Longrightarrow \text{biais !} \end{cases}$

Biais : estimation ponctuelle d'une variance (2/5)

- X : variable aléatoire sur l'ensemble de la population
- σ^2 : variance de X
- échantillon i.i.d.de taille n
- X_i : variable aléatoire correspondant au $i$ème individu
- x_i : valeur observée de X_i
- $\overline{X}$: variable aléatoire « moyenne sur l'échantillon »
- $\overline{x}$: valeur observée de $\overline{X}$
- $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - \overline{x}^2$
- s_n^2 = variance de l'échantillon

Problème : pourquoi s_n^2 n'est-il pas un bon estimateur de σ^2 ?

Biais : estimation ponctuelle d'une variance (3/5)

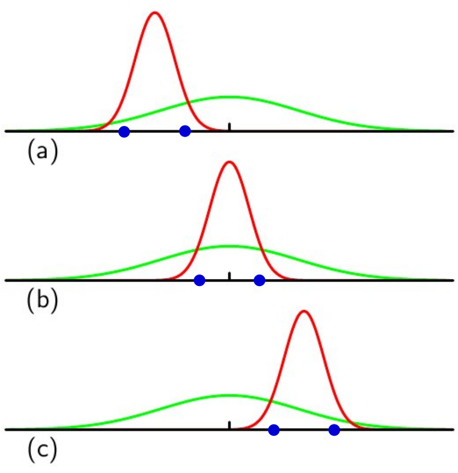
- $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2\right) - \overline{X}^2$
- s_n^2 est une réalisation de S_n^2
- $E(S_n^2) = E\left(\frac{1}{n}\left(\sum_{i=1}^{n} X_i^2\right) - \overline{X}^2\right) = \frac{1}{n}\left(\sum_{i=1}^{n} E(X_i^2)\right) - E\left(\overline{X}^2\right)$
- Or échantillon i.i.d $\Longrightarrow \forall i,\ E(X_i^2) = E(X^2)$

  $$\Longrightarrow E(S_n^2) = E(X^2) - E(\overline{X}^2)$$
- Or $V(X) = E(X^2) - E(X)^2$ et $V(\overline{X}) = E(\overline{X}^2) - E(\overline{X})^2 = E(\overline{X}^2) - E(X)^2$

  $\Longrightarrow E(S_n^2) = V(X) + E(X)^2 - V(\overline{X}) - E(X)^2 = V(X) - V(\overline{X})$
- Or échantillon i.i.d $\Longrightarrow V(X) = \sigma^2$ et $V(\overline{X}) = \frac{\sigma^2}{n}$

$$\Longrightarrow E(S_n^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

Biais : estimation ponctuelle d'une variance (4/5)



- courbe verte : la population
- courbes rouges : échantillons
- pts bleus : valeurs observées
- variance S_n^2 sous-estimée : mesurée par rapport à la moyenne de l'échantillon au lieu de la moyenne de la population

Biais : estimation ponctuelle d'une variance (5/5)

Variance corrigée

- X : variable aléatoire sur la population
- σ^2 : variance de X sur cette population
- échantillon de taille n avec remise
- X_i : variable aléatoire correspondant au $i$ème individu
- $\overline{X}$: moyenne des X_i
- alors : $E\left(\frac{n}{n-1}\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{n}\right) = \sigma^2$
- *Variance corrigée :* $\frac{n}{n-1}$ fois la variance de l'échantillon S_n^2

RFIDEC — cours 3 : Intervalles de confiance, tests d'hypothèses, loi du χ^2

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours n˚3

1. Intervalles de confiance
2. Tests d'hypothèses
3. La loi du χ^2

Intervalles de confiance

Estimateur T d'un paramètre $\theta \Longrightarrow$ valeur estimée $\hat{\theta}$

Problème : peut-on avoir confiance dans l'estimation ponctuelle ?

Intervalle de confiance

Un *intervalle de confiance* de niveau $1 - \alpha$ = intervalle

$]a(T), b(T)[$ tel que :

$$\forall \theta \in \Theta,\ P_\theta(]a(T), b(T)[\ni \theta) = 1 - \alpha$$

⚠ $1 - \alpha$ = proba que l'intervalle contienne θ

Intervalles de confiance : exemple (1/2)

$X \sim \mathcal{N}(\mu; \sigma^2)$

échantillon de taille $n \Longrightarrow \overline{X}$ = moyenne

théorème central-limite $\Longrightarrow \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0; 1)$

$$\begin{cases} n \text{ très grand} \Longrightarrow \text{la valeur observée } \overline{x} \text{ de } \overline{X} \approx \mu \\ n \text{ moins grand} \Longrightarrow \overline{x} \not\approx \mu \end{cases}$$

$\Longrightarrow$ estimation par intervalle de confiance de niveau 95%

loi normale $\Longrightarrow P\left(-1,96 \leq \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq 1,96\right) = 95\%$

Intervalles de confiance : exemple (2/2)

$$P\left(-1,96 \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq 1,96\right) = 95\%$$

$$P\left(\overline{X} - 1,96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1,96\frac{\sigma}{\sqrt{n}}\right) = 95\%$$

$\Longrightarrow$ intervalle de confiance = $\left]\overline{x} - 1,96\frac{\sigma}{\sqrt{n}}, \overline{x} + 1,96\frac{\sigma}{\sqrt{n}}\right[$

⚠ seulement maintenant, on tire un échantillon de taille n

$\Longrightarrow$ observation de $\overline{x}$

$\Longrightarrow$ on peut calculer $\left]\overline{x} - 1,96\frac{\sigma}{\sqrt{n}}, \overline{x} + 1,96\frac{\sigma}{\sqrt{n}}\right[$

Intervalles de confiance : autre exemple (1/2)

Énoncé de l'exemple

- plusieurs centaines de candidats à un examen
- variance sur les notes obtenues ≈ 16
- correcteur $\Longrightarrow$ noté 100 copies, moyenne = 8,75
- Problème : moyenne sur toutes les copies de l'examen ?
- hypothèse : les notes suivent une loi normale $\mathcal{N}(\mu; 16)$

$\overline{X}$ = variable aléatoire ≪ moyenne des notes d'un correcteur ≫

théorème central-limite $\Longrightarrow \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \dfrac{\overline{X} - \mu}{4/10} \sim \mathcal{N}(0; 1)$

chercher dans la table de la loi normale $z_{\alpha/2}$ tel que :

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

Intervalles de confiance : autre exemple (2/2)

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

$1-\alpha$	intervalle de confiance		
50%	$[8,75 - 0,674 \times 0,4 ; 8,75 + 0,674 \times 0,4]$	$=$	$[8,48 ; 9,02]$
75%	$[8,75 - 1,15 \times 0,4 ; 8,75 + 1,15 \times 0,4]$	$=$	$[8,29 ; 9,21]$
80%	$[8,75 - 1,28 \times 0,4 ; 8,75 + 1,28 \times 0,4]$	$=$	$[8,24 ; 9,26]$
90%	$[8,75 - 1,645 \times 0,4 ; 8,75 + 1,645 \times 0,4]$	$=$	$[8,09 ; 9,41]$
95%	$[8,75 - 1,96 \times 0,4 ; 8,75 + 1,96 \times 0,4]$	$=$	$[7,96 ; 9,53]$
99%	$[8,75 - 2,575 \times 0,4 ; 8,75 + 2,575 \times 0,4]$	$=$	$[7,72 ; 9,78]$

Exemple : analyse des déchets (cf. cours 2)

- Grenelle de l'environnement
  - $\Longrightarrow$ réduction des déchets
  - $\Longrightarrow$ analyse des déchets
  - $\Longrightarrow$ échantillon de taille 100
- $\overline{x}$: moyenne de l'échantillon $=$ 390 kg/an/habitant
- écart-type $\sigma = 20$ supposé connu
- $\overline{X} \sim \mathcal{N}(\mu, 4)$

$\Longrightarrow P\left(-Z_{\alpha/2} \leq \dfrac{\overline{X}-\mu}{2} \leq Z_{\alpha/2}\right) = 1-\alpha$

$\Longrightarrow$ estimation par intervalle de confiance de niveau $1-\alpha$:

$]\overline{x} - 2z_{\alpha/2}, \overline{x} + 2z_{\alpha/2}[$

Exemple : analyse des déchets (suite)

Estimation par intervalle de confiance de niveau $1-\alpha$:
$]\overline{x} - 2z_{\alpha/2}, \overline{x} + 2z_{\alpha/2}[$

$1-\alpha$	intervalle de confiance		
50%	$[390 - 0,674 \times 2 ; 390 + 0,674 \times 2]$	$=$	$[388,65 ; 391,35]$
75%	$[390 - 1,15 \times 2 ; 390 + 1,15 \times 2]$	$=$	$[387,70 ; 392,30]$
80%	$[390 - 1,28 \times 2 ; 390 + 1,28 \times 2]$	$=$	$[387,44 ; 392,56]$
90%	$[390 - 1,645 \times 2 ; 390 + 1,645 \times 2]$	$=$	$[386,71 ; 393,29]$
95%	$[390 - 1,96 \times 2 ; 390 + 1,96 \times 2]$	$=$	$[386,08 ; 393,92]$
99%	$[390 - 2,575 \times 2 ; 390 + 2,575 \times 2]$	$=$	$[384,85 ; 395,15]$

Exemple du réchauffement climatique (cf. cours 2)

- opinion des gens sur le réchauffement climatique
- 1000 personnes de 15 ans et $+$ interrogées
- 790 pensent qu'il y a un changement climatique
- 210 ne le pensent pas
- $\overline{P}$: proportion de succès moyenne de l'échantillon
- p : proportion de personnes pensant qu'il y a dérèglement climatique dans la population française

$$\frac{\overline{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0;1)$$

$\Longrightarrow$ estimation par intervalle de confiance de niveau $1 - \alpha$:

$$\left]\overline{p} - \sqrt{\tfrac{p(1-p)}{n}}z_{\alpha/2}; \overline{p} + \sqrt{\tfrac{p(1-p)}{n}}z_{\alpha/2}\right[\approx \left]\overline{p} - \sqrt{\tfrac{\overline{p}(1-\overline{p})}{n}}z_{\alpha/2}; \overline{p} + \sqrt{\tfrac{\overline{p}(1-\overline{p})}{n}}z_{\alpha/2}\right[$$

Tests d'hypothèses en statistique classique (1/2)

Hypothèses

- Θ = ensemble des valeurs du paramètre θ
- Θ partitionné en Θ_0 et Θ_1
- *hypothèses* = assertions H_0 = "$\theta \in \Theta_0$" et H_1 = "$\theta \in \Theta_1$"
- H_0 = *hypothèse nulle*, H_1 = *contre-hypothèse*
- hypothèse H_i est *simple* si Θ_i est un singleton ; sinon elle est *multiple*
- test *unilatéral* = valeurs dans Θ_1 toutes soit plus grandes, soit plus petites, que celles dans Θ_0 ; sinon test *bilatéral*

Tests d'hypothèses en statistique classique (2/2)

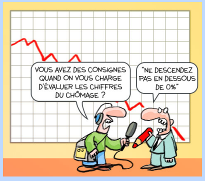
	hypothèse	test
$H_0 : \mu = 4$	simple	unilatéral
$H_1 : \mu = 6$	simple	
$H_0 : \mu = 4$	simple	test unilatéral
$H_1 : \mu > 4$	composée	
$H_0 : \mu = 4$	simple	test bilatéral
$H_1 : \mu \neq 4$	composée	
$H_0 : \mu = 4$	simple	formulation incorrecte : les hypothèses
$H_1 : \mu > 3$	composée	ne sont pas mutuellement exclusives

Exemples pratiques d'hypothèses

- association de consommateurs
- échantillon de 100 bouteilles de Bordeaux
- *Pb :* la quantité de vin est-elle bien égale à 75cl ?

- paramètre θ étudié = $\mu = E(X)$
- X = quantité de vin dans les bouteilles
- rôle de l'association $\Longrightarrow$ $H_0 : \mu = 75\text{cl}$ et $H_1 : \mu < 75\text{cl}$

- le mois dernier, taux de chômage $= 10\%$
- échantillon : 400 individus de la pop. active
- *Pb :* le taux de chômage a-t-il été modifié ?



- paramètre étudié = p = % de chômeurs
- $H_0 : p = 10\%$ et $H_1 : p \neq 10\%$

Tests d'hypothèse

Définition du test

- *test* entre deux hypothèses H_0 et H_1 = *règle de décision* δ
- règle fondée sur les observations
- ensemble des décisions possibles = $\mathcal{D} = \{d_0, d_1\}$
- $d_0 =$ "*accepter* H_0"
- $d_1 =$ "*accepter* H_1" $=$ "*rejeter* H_0"

région critique

- échantillon $\Longrightarrow$ n-uplet $(x_1, \dots, x_n)$ de valeurs (dans $\mathbb{R}$)
- δ = fonction $\mathbb{R}^n \mapsto \mathcal{D}$
- *région critique :* $W = \{n\text{-uplets } \mathbf{x} \in \mathbb{R}^n : \delta(\mathbf{x}) = d_1\}$
- région critique = *région de rejet*
- *région d'acceptation* = $A = \{\mathbf{x} \in \mathbb{R}^n : \delta(\mathbf{x}) = d_0\}$

Régions critiques

Hypothèses	Règle de décision
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	« rejeter H_0 si $\overline{x} > c$ », où c est un nombre plus grand que μ_0
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	« rejeter H_0 si $\overline{x} < c$ », où c est un nombre plus petit que μ_0
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	« rejeter H_0 si $\overline{x} < c_1$ ou $c_2 < \overline{x}$ », où c_1 et c_2 sont des nombres respectivement plus petit et plus grand que μ_0, et également éloignés de celui-ci

Problème : erreurs dans les décisions prises

Erreurs dans les décisions

Décision prise \ Réalité	H_0 est vraie	H_1 est vraie
H_0 est rejetée	mauvaise décision : erreur de type I	bonne décision
H_0 n'est pas rejetée	bonne décision	mauvaise décision : erreur de type II

$$\begin{aligned}\alpha &= \text{risque de première espèce}\\ &= \text{probabilité de réaliser une erreur de type I}\\ &= \text{probabilité de rejeter } H_0 \text{ sachant que } H_0 \text{ est vraie}\\ &= P(\text{rejeter } H_0 | H_0 \text{ est vraie}),\end{aligned}$$

$$\begin{aligned}\beta &= \text{risque de deuxième espèce}\\ &= \text{probabilité de réaliser une erreur de type II}\\ &= \text{probabilité de rejeter } H_1 \text{ sachant que } H_1 \text{ est vraie}\\ &= P(\text{rejeter } H_1 | H_1 \text{ est vraie}).\end{aligned}$$

Exemple de calcul de α (1/2)

Exemple

- échantillon de taille 25
- paramètre estimé : μ d'une variable $X \sim \mathcal{N}(\mu; 100)$
- hypothèses : $H_0 : \mu = 10 \quad H_1 : \mu > 10$

Sous H_0 : $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \dfrac{\overline{X} - 10}{10/5} = \dfrac{\overline{X} - 10}{2} \sim \mathcal{N}(0; 1)$

Sous H_0 : peu probable que $\overline{X}$ éloignée de plus de 2 écarts-types de μ (4,56% de chance)

$\Longrightarrow$ peu probable que $\overline{X} < 6$ ou $\overline{X} > 14$

$\Longrightarrow$ région critique pourrait être « rejeter H_0 si $\overline{x} > 14$ »

Exemple de calcul de α (2/2)

- échantillon de taille 25
- paramètre estimé : μ d'une variable $X \sim \mathcal{N}(\mu; 100)$
- hypothèses : $H_0 : \mu = 10 \quad H_1 : \mu > 10$
- région critique : « rejeter H_0 si $\overline{x} > 14$ »

$$\begin{aligned}\alpha &= P(\text{rejeter } H_0 | H_0 \text{ est vraie})\\ &= P(\overline{X} > 14 | \mu = 10)\\ &= P\left(\frac{\overline{X} - 10}{2} > \frac{14 - 10}{2} \middle| \mu = 10\right)\\ &= P\left(\frac{\overline{X} - 10}{2} > 2\right) = 0,0228\end{aligned}$$

⚠ en principe α est fixé et on cherche la région critique

Exemple de test d'hypothèses (1/2)

- filtre de mails sur un serveur mail :

mail → extraction de caractéristiques → agrégation $\Longrightarrow$ score → filtrage par score

- X = score $\geq 18000 \Longrightarrow$ spam ; historiques des mails $\Longrightarrow \sigma_X = 5000$
- le serveur reçoit un envoi en masse de $n = 400$ mails de xx@yy.fr

  Problème : xx@yy.fr est-il un spammeur ?
- H_0 : xx@yy.fr = « spammeur » v.s. H_1 : xx@yy.fr $\neq$ « spammeur »
- test : $H_0 : \mu = 18000$ v.s. $H_1 : \mu < 18000$ où $\mu = E(X)$
- règle : si $\overline{x} < c$ alors rejeter H_0
- 400 mails $\Longrightarrow$ théorème central limite $\Longrightarrow$ sous H_0 :

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 18000}{5000/\sqrt{400}} = \frac{\overline{X} - 18000}{250} \sim \mathcal{N}(0;1)$$

Exemple de test d'hypothèses (2/2)

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 18000}{5000/\sqrt{400}} = \frac{\overline{X} - 18000}{250} \sim \mathcal{N}(0;1)$$

- choix du risque de première espèce : $\alpha = 0,01$
- $\alpha = 0,01 = P(\overline{X} < c | \mu = 18000)$

$$= P\left(\frac{\overline{X} - 18000}{250} < \frac{c - 18000}{250} | \mu = 18000\right)$$

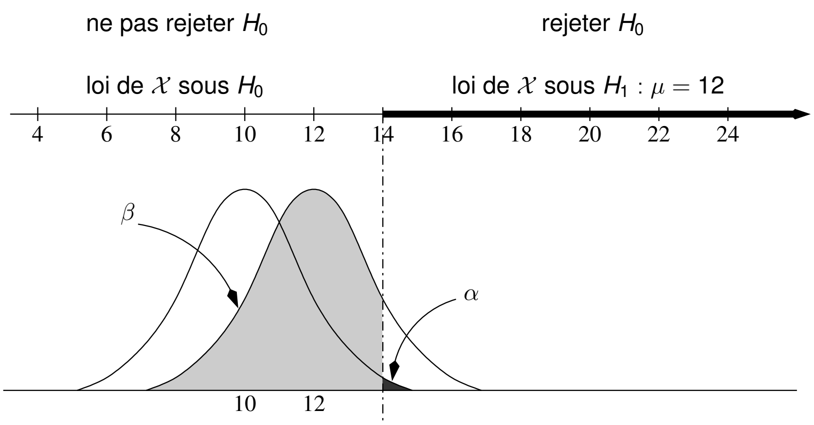
$$= P\left(Z < \frac{c - 18000}{250}\right)$$

$$= P(Z < -2,326)$$

$$\Longrightarrow \frac{c - 18000}{250} = -2,326 \Longrightarrow c = 17418,5$$

règle de décision : si $\overline{x} < 17418,5$, rejeter $H_0 \Longrightarrow$ non spam

Interprétation de α et β



Puissance du test

$$\alpha = P(\text{rejeter } H_0 | H_0 \text{ est vraie})$$
$$\beta = P(\text{rejeter } H_1 | H_1 \text{ est vraie})$$

α et β varient en sens inverse l'un de l'autre

$\Longrightarrow$ test = compromis entre les deux risques

H_0 = hypothèse privilégiée, vérifiée jusqu'à présent et que l'on n'aimerait pas abandonner à tort

$\Longrightarrow$ on fixe un *seuil* α_0 :

- $\alpha \leq \alpha_0$
- test minimisant β sous cette contrainte
- min β = max $1 - \beta$

$1 - \beta$ = puissance du test

Exemple de calcul de β (1/2)

- échantillon de taille 25
- paramètre estimé : μ d'une variable $X \sim \mathcal{N}(\mu; 100)$
- hypothèses : $H_0 : \mu = 10 \quad H_1 : \mu > 10$
- région critique : « rejeter H_0 si $\overline{x} > 14$ »

sous H_1 : plusieurs valeurs de μ sont possibles

$\Longrightarrow$ courbe de puissance du test en fonction de μ

Supposons que $\mu = 11$:

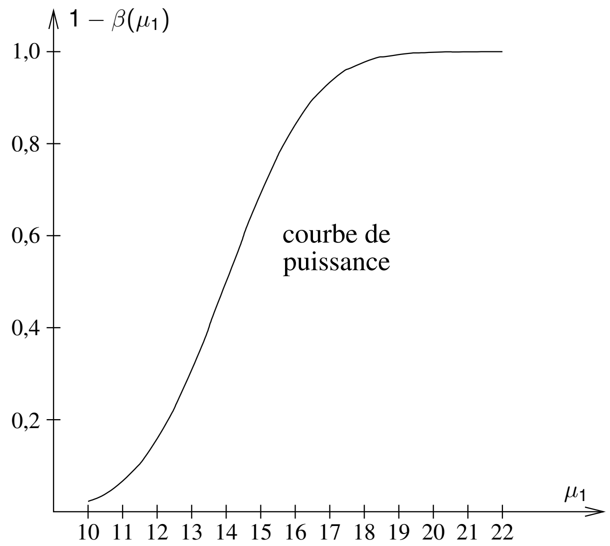
$$\mu = 11 \Longrightarrow \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 11}{2} \sim \mathcal{N}(0; 1)$$

Exemple de calcul de β (2/2)

$$\begin{aligned} 1 - \beta(11) &= P(\text{rejeter } H_0 | H_1 : \mu = 11 \text{ est vraie}) \\ &= P(\overline{X} > 14 | \mu = 11) \\ &= P\left(\frac{\overline{X} - 11}{2} > \frac{14 - 11}{2} | \mu = 11\right) \\ &= P\left(\frac{\overline{X} - 11}{2} > 1,5\right) = 0,0668 \end{aligned}$$

μ_1	$z_1 = \frac{14 - \mu_1}{2}$	$1 - \beta(\mu_1) = P(Z > z_1)$	$\beta(\mu_1)$
10	2,0	0,0228	0,9772
11	1,5	0,0668	0,9332
12	1,0	0,1587	0,8413
13	0,5	0,3085	0,6915
14	0,0	0,5000	0,5000
15	-0,5	0,6915	0,3085
16	-1,0	0,8413	0,1587
17	-1,5	0,9332	0,0668

Courbe de puissance du test



Exemple : notes d'examen de RFIDEC (1/3)

- les années précédentes, notes d'examen $\sim \mathcal{N}(14, 6^2)$
- cette année, correction d'un échantillon de 9 copies :

10	8	13	20	12	14	9	7	15

Les notes sont-elles en baisse cette année ?

- hypothèse $H_0 =$ « la moyenne est égale à 14 »

  hypothèse $H_1 =$ « la moyenne a baissé, i.e., elle est ≤ 14 »

  test d'hypothèse de niveau de confiance $1-\alpha = 95\%$

$\Longrightarrow$ déterminer seuil c tel que $\overline{x} < c \Longrightarrow H_1$ plus probable que H_0

Exemple : notes d'examen de RFIDEC (2/3)

10	8	13	20	12	14	9	7	15

$H_0 : \mu = 14, \sigma = 6$

- sous hypothèse H_0, on sait que $\dfrac{\overline{X}-14}{\sigma/\sqrt{n}} = \dfrac{\overline{X}-14}{2} \sim \mathcal{N}(0;1)$
- calcul du seuil c (région de rejet) :

$$P\left(\frac{\overline{X}-14}{2} < \frac{c-14}{2} \,\middle|\, \frac{\overline{X}-14}{2} \sim \mathcal{N}(0;1)\right) = 0,05$$

- Table de la loi normale : $\frac{c-14}{2} \approx -1,645 \Longrightarrow c = 10,71$
- Règle de décision : rejeter H_0 si $\overline{x} < 10,71$
- tableau $\Longrightarrow \overline{x} = 12$

  $\Longrightarrow$ on ne peut déduire que la moyenne a diminué

Exemple : notes d'examen de RFIDEC (3/3)

Problème : le risque de 2ème espèce est-il élevé ?

Puissance du test pour une moyenne de 12

- H_1 : la moyenne est égale à 12
- Puissance du test $= 1 - \beta(12)$

$$= P(\text{rejeter } H_0 | H_1)$$

$$= P\left(\overline{X} < 10,71 \,\middle|\, \frac{\overline{X}-12}{2} \sim \mathcal{N}(0;1)\right)$$

$$= P\left(\frac{\overline{X}-12}{2} < -0,645 \,\middle|\, \frac{\overline{X}-12}{2} \sim \mathcal{N}(0;1)\right)$$

$$\approx 25,95\%.$$

Lemme de Neyman-Pearson (1/2)

Cas : $\Theta_0 = \{\theta_0\}$ $\quad\Theta_1 = \{\theta_1\}$

- Échantillon $(x_1, \dots, x_n)$ de taille n
- Échantillon $\Longrightarrow$ les x_i = réalisations de variables aléatoires X_i
- Échantillon i.i.d. $\Longrightarrow$ les X_i sont mutuellement indépendants

$\Longrightarrow P(X_1 = x_1, \dots, X_n = x_n | \theta = \theta_k) = \prod_{i=1}^n P(X_i = x_i | \theta = \theta_k)$

Vraisemblance d'un échantillon

- $x = (x_1, \dots, x_n)$: échantillon de taille n
- $L(x, \theta_k)$ = Vraisemblance de l' échantillon
- $L(x, \theta_k)$ = proba d'obtenir cet échantillon sachant que $\theta = \theta_k$

$$L(x, \theta_k) = P(x_1, \dots, x_n | \theta = \theta_k) = \prod_{i=1}^n P(x_i | \theta = \theta_k)$$

Lemme de Neyman-Pearson (2/2)

Cas : $\Theta_0 = \{\theta_0\}$ $\quad\Theta_1 = \{\theta_1\}$

Lemme de Neyman-Pearson

- il existe toujours un test (aléatoire) le plus puissant de seuil donné α_0
- c'est un test du rapport de vraisemblance :

$$\frac{L(x,\theta_0)}{L(x,\theta_1)} > k \Rightarrow x \in A \text{ (accepter } H_0)$$

$$\frac{L(x,\theta_0)}{L(x,\theta_1)} < k \Rightarrow x \in W \text{ (rejeter } H_0)$$

$$\frac{L(x,\theta_0)}{L(x,\theta_1)} = k \Rightarrow \delta(x) = \rho \text{ (accepter } H_0 \text{ avec proba } 1 - \rho$$

$$H_1 \text{ avec proba } \rho)$$

- k et ρ déterminés de façon unique par $\alpha = \alpha_0$

Loi du χ^2 (1/3)

- population $\Longrightarrow$ répartie en k classes

p_1	p_2	p_3		p_k

- hypothèse : répartition dans les classes connues

  $\Longrightarrow p_r =$ proba qu'un individu appartienne à la classe c_r
- échantillon de n individus
- N_r = variable aléatoire « nombre d'individus tirés de classe c_r »
- Chaque individu $\Longrightarrow p_r$ chances d'appartenir à la classe c_r

  $\Longrightarrow X_i^r =$ v.a. succès si l'individu i appartient à la classe c_r

  $\Longrightarrow X_i^r \sim \mathcal{B}(1, p_r)$

  $\Longrightarrow N_r \sim \mathcal{B}(n, p_r)$

  $\Longrightarrow N_r \sim$ loi normale quand n grand

Loi du χ^2 (2/3)

- population $\Longrightarrow$ répartie en k classes

p_1	p_2	p_3		p_k

- $p_r =$ proba qu'un individu appartienne à la classe c_r
- échantillon de n individus
- N_r = v.a. « nb d'individus tirés de classe c_r » $\sim$ loi normale

$$D^2_{(n)} = \sum_{r=1}^{k} \frac{(N_r - n.p_r)^2}{n.p_r}$$

  $\Longrightarrow D^2_{(n)} =$ somme des carrés de k v.a. $\sim$ lois normales
- $D^2_{(n)}$ = écart entre théorie et observation
- $D^2_{(n)}$ tend en loi, lorsque $n \to \infty$, vers une loi du χ^2_{k-1}

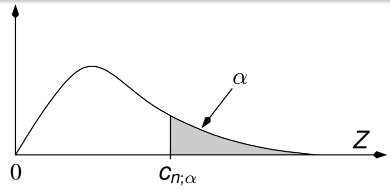
Loi du χ^2 (3/3)

Loi du χ^2

- loi du χ^2_r = la loi de la somme des carrés de r variables indépendantes et de même loi $\mathcal{N}(0, 1)$
- espérance = r
- variance = $2r$

Table de la loi du χ^2

valeurs dans le tableau
ci-dessous : les $c_{n;\alpha}$
tels que $P(Z > c_{n;\alpha}) = \alpha$



$n \setminus \alpha$	0,995	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,005
1	0,00004	0,0002	0,001	0,0039	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,103	0,211	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	16,0	18,3	20,5	23,2	25,2

RFIDEC — cours 4 : Tests d'ajustement, apprentissage de paramètres

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours n˚4

1. Tests d'ajustement
2. Tests d'indépendance
3. Application aux réseaux bayésiens
4. Maximum de vraisemblance

Tests d'ajustement

Définition

- test d'ajustement = test $\Longrightarrow$ 2 issues possibles :
  1. acceptation de l'hypothèse que l'échantillon observé est tiré selon une certaine loi
  2. rejet de l'hypothèse 1
- contre-hypothèse : ne précise pas de quelle autre loi l'échantillon aurait pu être tiré

Tests d'ajustement II : le retour du χ^2

- population répartie en k classes
- échantillon de taille $n \Longrightarrow$ répartition = $(n_1, \dots, n_k)$
- supposons l'échantillon tiré selon la loi multinomiale $(p_1, \dots, p_k)$

  $\Longrightarrow (n_1, \dots, n_k) \approx (n.p_1, \dots, n.p_k)$

Rappel : $D^2_{(n)} = \sum_{r=1}^{k} \frac{(N_r - n.p_r)^2}{n.p_r} \sim \chi^2_{k-1}$

- d^2 valeur prise par $D^2_{(n)}$

  $\Longrightarrow$ si échantillon tiré selon $(p_1, \dots, p_k)$ alors d^2 petit
- table de la loi du $\chi^2 \Longrightarrow d^2_\alpha$ tel que $P(\chi^2_{k-1} > d^2_\alpha) = \alpha$

  $\Longrightarrow$ règle de décision : si $d^2 < d^2_\alpha$ alors OK

Tests d'ajustement en pratique

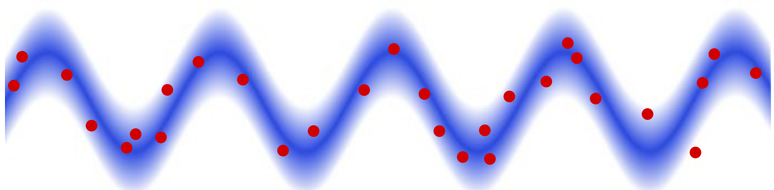
Mise en place d'un test d'ajustement

1. population répartie en k classes
2. échantillon de taille $n \Longrightarrow$ répartition $= (n_1, \dots, n_k)$
3. on vérifie si l'échantillon tiré selon la loi $(p_1, \dots, p_k)$:
   - A. choix du risque de première espèce α
   - B. calcul de $d^2 = \sum_{r=1}^{k} \frac{(n_r - n.p_r)^2}{n.p_r}$
   - C. lecture dans une table de d^2_α tel que $P(\chi^2_{k-1} > d^2_\alpha) = \alpha$
   - D. si $d^2 < d^2_\alpha$ alors règle de décision :

     $(p_1, \dots, p_k)$ est la loi selon laquelle est tiré l'échantillon

     sinon l'échantillon est tiré selon une autre loi

Exemple de test d'ajustement (1/3)



- observations = ● = $\{(x_i, y_i)\}$
- *Problème :* les ● proviennent-ils de points situés sur la courbe $y = sin(x)$ mais observés avec un bruit gaussien ?

$\Longrightarrow$ problème : $T_i = Y_i - \sin(x_i) \sim \mathcal{N}(0, 1)$?

observations des t_i, réparties en 8 classes :

t_i	$]-\infty; -3[$	$[-3; -2[$	$[-2; -1[$	$[-1; 0[$	$[0; 1[$	$[1; 2[$	$[2; 3[$	$[3; +\infty[$
N_r	1	2	13	35	30	15	3	1

Exemple de test d'ajustement (2/3)

Rappel : $T_i \sim \mathcal{N}(0,1)$

t_i	$]-\infty;-3[$	$[-3;-2[$	$[-2;-1[$	$[-1;0[$	$[0;1[$	$[1;2[$	$[2;3[$	$[3;+\infty[$
N_r	1	2	13	35	30	15	3	1
$n.p_r$	0.14	2.14	13.59	34.13	34.13	13.59	2.14	0.14

$$\Longrightarrow d^2 = \sum_{r=1}^{8} \frac{(n_r - n.p_r)^2}{n.p_r} \approx 11.61$$

pour $\alpha = 0.05$, $P(\chi^2_7 > d^2_\alpha) = \alpha \Longrightarrow d^2_\alpha = 14.1$

$\Longrightarrow d^2 < d^2_\alpha \Longrightarrow$ règle de décision :

l'échantillon est bien tiré selon $sin(x)+$ un bruit gaussien

Exemple de test d'ajustement (3/3)

Nouvel échantillon :

t_i	$]-\infty;-3[$	$[-3;-2[$	$[-2;-1[$	$[-1;0[$	$[0;1[$	$[1;2[$	$[2;3[$	$[3;+\infty[$
N_r	2	2	12	35	30	15	3	1
$n.p_r$	0.14	2.14	13.59	34.13	34.13	13.59	2.14	0.14

$$\Longrightarrow d^2 = \sum_{r=1}^{8} \frac{(n_r - n.p_r)^2}{n.p_r} \approx 31.20$$

pour $\alpha = 0.05$, $P(\chi^2_7 > d^2_\alpha) = \alpha \Longrightarrow d^2_\alpha = 14.1$

$\Longrightarrow d^2 > d^2_\alpha \Longrightarrow$ règle de décision :

l'échantillon n'est pas tiré selon $sin(x)+$ un bruit gaussien

Exemple de test d'ajustement (1/2)

- péage d'autoroute : 10 cabines
- nombre de clients / cabine sur une heure :

N˚ cabine	1	2	3	4	5	6	7	8	9	10
Nb clients	24	14	18	20	23	13	23	24	23	18

Clients distribués uniformément sur l'ensemble des cabines ?

$\Longrightarrow$ test d'ajustement, niveau de confiance : $1 - \alpha = 95\%$

- H_0 = « la répartition des clients est uniforme »

  H_1 = « la répartition n'est pas uniforme »
- $H_0 \Longrightarrow$ 20 clients / cabine (uniforme)

Exemple de test d'ajustement (2/2)

- X_i : variable « effectif » recensé pour la $i$ème cabine
- Statistique d'ajustement : $D^2 = \sum_{i=1}^{10} \frac{(X_i-20)^2}{20}$
- $D^2 \sim \chi^2_9$
- $\alpha = 0,05 = P(\text{rejeter } H_0 | H_0 \text{ est vraie})$

  $= P\left(D^2 > d_\alpha \mid D^2 \sim \chi^2_9\right)$

  $\Longrightarrow d_\alpha = 16,9$
- calcul de la valeur de d observée sur l'échantillon :

$$d^2 = \frac{1}{20}[(14-20)^2 + (24-20)^2 + (18-20)^2 + (20-20)^2 + (23-20)^2 + (13-20)^2 + (23-20)^2 + (18-20)^2 + (24-20)^2 + (23-20)^2] = 7,6.$$

$\Longrightarrow$ estimation : répartition uniforme

Tests d'indépendance (1/3)

- 2 caractères X et Y
- classes de X : $A_1, A_2, \ldots, A_I$
- classes de Y : $B_1, B_2, \ldots, B_J$
- échantillon de taille n
- tableau de contingence :

$X \backslash Y$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}

Tests d'indépendance (2/3)

$X \backslash Y$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	total
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\cdot}$
total	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot J}$	n

$$\frac{n_{ij}}{n} = P(X \in A_i, Y \in B_j)$$

$$P(X \in A_i) = \frac{n_{i\cdot}}{n} = \frac{\sum_{j=1}^{J} n_{ij}}{n} \quad \text{et} \quad P(Y \in B_j) = \frac{n_{\cdot j}}{n} = \frac{\sum_{i=1}^{I} n_{ij}}{n}$$

X et Y indépendants $\Longrightarrow P(X \in A_i, Y \in B_j) = P(X \in A_i) \times P(Y \in B_j)$

Tests d'indépendance (3/3)

$X\backslash Y$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	*total*
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1.}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I.}$
total	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.j}$	$\cdots$	$n_{.J}$	n

$$X \text{ et } Y \text{ indépendants} \Longrightarrow \frac{n_{ij}}{n} = \frac{n_{i.}}{n} \times \frac{n_{.j}}{n} \Longrightarrow n_{ij} = \frac{n_{i.} \times n_{.j}}{n}$$

$$\chi^2_{(I-1)\times(J-1)} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}}$$

Exemple de test d'indépendance (1/2)

- notes d'examen de RFIDEC $\Longrightarrow$ 3 classes :

c_1	c_2	c_3
note < 8	note $\in [8, 12[$	note ≥ 12

- X : variable aléatoire « note 1ère session »
- Y : variable aléatoire « note 2ème session »

X et Y sont-elles des variables aléatoires indépendantes ?

- sélection d'un échantillon de 100 notes :

$X\backslash Y$	c_1	c_2	c_3
c_1	2	13	6
c_2	11	27	13
c_3	3	17	8

Exemple de test d'indépendance (2/2)

Test d'indépendance de niveau de confiance 90%

1. calcul des marginales :

$X\backslash Y$	c_1	c_2	c_3	total
c_1	2	13	6	21
c_2	11	27	13	51
c_3	3	17	8	28
total	16	57	27	

2. tableau obtenu si X et Y sont indépendants :

$X\backslash Y$	c_1	c_2	c_3
c_1	3.36	11.97	5.67
c_2	8.16	29.07	13.77
c_3	4.48	15.96	7.56

3. calcul de la statistique d^2 : $d^2 = 2,42$
4. $D^2 \sim \chi^2_4 \Longrightarrow d^2_\alpha = 7,78 \Longrightarrow d^2 < d^2_\alpha \Longrightarrow$ indépendance

Rappel sur l'indépendance conditionnelle

Indépendance conditionnelle de deux variables discrètes

X et Y sont *indépendantes* conditionnellement à Z si :

- $P(X \cap Y|Z) = P(X|Z) \times P(Y|Z)$
- si $P(Y|Z) > 0$ alors $P(X|Y,Z) = P(X|Z)$
- si $P(X|Z) > 0$ alors $P(Y|X,Z) = P(Y|Z)$

Interprétation

- Conditionnement = apport de connaissances
- Si l'on connaît la valeur de la variable Z, alors connaître celle de Y n'apporte rien sur la connaissance de X

Ces formules s'étendent si X, Y et/ou Z sont remplacés par des ensembles de variables aléatoires disjoints 2 à 2

Application en informatique

- n variables aléatoires $X_1, \ldots, X_n$
- $P(X_n, \ldots, X_1) = P(X_n|X_{n-1}, \ldots, X_1)P(X_{n-1}, \ldots, X_1)$
- Par récurrence :

$$P(X_n, \ldots, X_1) = P(X_1) \times \prod_{i=2}^{n} P(X_i|X_1, \ldots, X_{i-1})$$

- $\forall i$, $\{X_1, \ldots, X_{i-1}\} = L_i \cup K_i$, où $L_i \cap K_i = \emptyset$ et X_i indépendant de L_i conditionnellement à K_i
- Alors :

$$P(X_n, \ldots, X_1) = P(X_1) \times \prod_{i=2}^{n} P(X_i|K_i)$$

- Tables de proba $P(X_i|K_i)$ plus petites que $P(X_i|X_1, \ldots, X_{i-1})$

Exemple d'application (1/5)

Exemple de la dyspnée (Lauritzen & Spiegelhalter (88))

La dyspnée peut être engendrée par une tuberculose, un cancer des poumons, une bronchite, par plusieurs de ces maladies, ou bien par aucune.
Un séjour récent en Asie augmente les chances de tuberculose, tandis que fumer augmente les risques de cancer des poumons.
Des rayons X permettent de détecter une tuberculose ou un cancer.
Un patient éprouve des difficultés à respirer. Dans quelle mesure peut-on dire qu'il est atteint de dyspnée ?

Variables aléatoires :

- D : dyspnée : oui/non
- C : cancer : oui/non
- A : Asie : oui/non
- R : rayons X : positif/négatif
- T : tuberculose : oui/non
- B : bronchite : oui/non
- F : fumer : oui/non

Exemple d'application (2/5)

Exemple de la dyspnée (Lauritzen & Spiegelhalter (88))

La dyspnée peut être engendrée par une tuberculose, un cancer des poumons, une bronchite, par plusieurs de ces maladies, ou bien par aucune. Un séjour récent en Asie augmente les chances de tuberculose, tandis que fumer augmente les risques de cancer des poumons. Des rayons X permettent de détecter une tuberculose ou un cancer. Un patient éprouve des difficultés à respirer. Dans quelle mesure peut-on dire qu'il est atteint de dyspnée ?

$$P(D, R, T, C, B, A, F) = P(D|R, T, C, B, A, F) \times P(R, T, C, B, A, F)$$

Or $P(D|R, T, C, B, A, F) = P(D|T, C, B)$

$$\Longrightarrow P(D, R, T, C, B, A, F) = P(D|T, C, B) \times P(R, T, C, B, A, F)$$

Exemple d'application (3/5)

Exemple de la dyspnée (Lauritzen & Spiegelhalter (88))

La dyspnée peut être engendrée par une tuberculose, un cancer des poumons, une bronchite, par plusieurs de ces maladies, ou bien par aucune. Un séjour récent en Asie augmente les chances de tuberculose, tandis que fumer augmente les risques de cancer des poumons. Des rayons X permettent de détecter une tuberculose ou un cancer. Un patient éprouve des difficultés à respirer. Dans quelle mesure peut-on dire qu'il est atteint de dyspnée ?

$$P(D, R, T, C, B, A, F) = P(D|T, C, B) \times P(R, T, C, B, A, F)$$

or $P(R, T, C, B, A, F) = P(R|T, C, B, A, F) \times P(T, C, B, A, F)$

et $P(R|T, C, B, A, F) = P(R|T, C)$

$$\Longrightarrow P(D, R, T, C, B, A, F) = P(D|T, C, B) \times P(R|T, C) \times P(T, C, B, A, F)$$

Exemple d'application (4/5)

Exemple de la dyspnée (Lauritzen & Spiegelhalter (88))

La dyspnée peut être engendrée par une tuberculose, un cancer des poumons, une bronchite, par plusieurs de ces maladies, ou bien par aucune. Un séjour récent en Asie augmente les chances de tuberculose, tandis que fumer augmente les risques de cancer des poumons. Des rayons X permettent de détecter une tuberculose ou un cancer. Un patient éprouve des difficultés à respirer. Dans quelle mesure peut-on dire qu'il est atteint de dyspnée ?

$$P(D, R, T, C, B, A, F) = P(D|T, C, B) \times P(R|T, C) \times P(T, C, B, A, F)$$

$$P(T|C, B, A, F) = P(T|A)$$

$$P(D, R, T, C, B, A, F) = P(D|T, C, B) \times P(R|T, C) \times P(T|A) \times P(C, B, A, F)$$

....

$$= P(D|T, C, B) \times P(R|T, C) \times P(T|A) \times P(C|F) \times P(B|F) \times P(A) \times P(F)$$

Exemple d'application (5/5)

$$P(D,R,T,C,B,A,F) = P(D|T,C,B) \times P(R|T,C) \times P(T|A) \times P(C|F) \times P(B|F) \times P(A) \times P(F)$$

Si toutes les variables ont 10 valeurs possibles :

$P(D,R,T,C,B,A,F)$ nécessite une table de 10^7 éléments

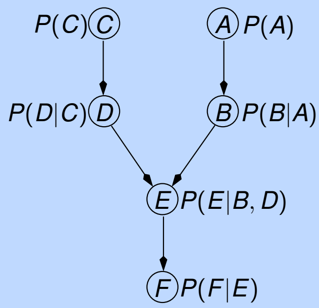
formule décomposée nécessite :

$10000 + 1000 + 3 \times 100 + 2 \times 10 = 11320$ éléments

Qu'est-ce qu'un réseau bayésien ?

Définition d'un réseau bayésien

1. un graphe sans circuit :



qui représente une décomposition de la loi jointe :
$P(A,B,C,D,E,F) = P(F|E)P(E|B,D)P(D|C)P(C)P(B|A)P(A)$

2. À chaque noeud X du graphe est associé sa probabilité conditionnellement à ses parents.

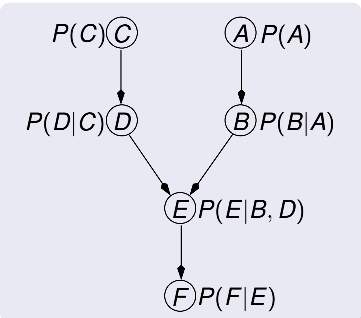
Quelques applications des réseaux bayésiens

diagnostic

- *diagnostic de panne*
- *sûreté de fonctionnement*
- *filtrage de spams*

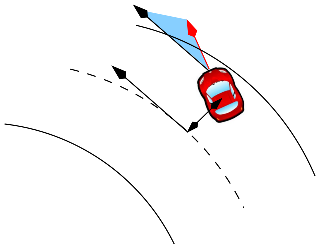
prédiction

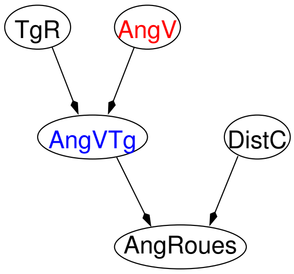
- *modélisation de joueurs*
- *prévisions boursières*



Diagnostic : une voiture bayésienne (1/3)

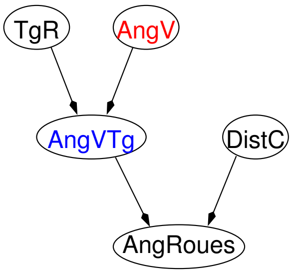
- RB conduit une voiture
- but : rester au milieu de la route





Problème : effectuer rapidement les calculs

Diagnostic : une voiture bayésienne (2/3)



décomposition :

$P(\text{AngRoues}|\text{AngVTg},\text{DistC})$

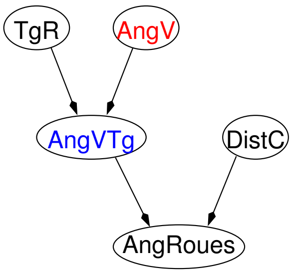
$P(\text{AngVTg}|\text{TgR},\text{AngV})$

$P(\text{TgR})P(\text{AngV})P(\text{DistC})$

Principe de fonctionnement de la voiture

- calculer $P(\text{AngRoues})$ puis tirer selon cette loi une action
- calculer Argmax $P(\text{AngRoues})$
- calculs $\Longrightarrow$ éliminer les variables une par une

Diagnostic : une voiture bayésienne (3/3)



décomposition :

$P(\text{AngRoues}|\text{AngVTg},\text{DistC})$

$P(\text{AngVTg}|\text{TgR},\text{AngV})$

$P(\text{TgR})P(\text{AngV})P(\text{DistC})$

Vraisemblance d'un échantillon : le cas discret

- Paramètre à estimer : θ
- Échantillon $\mathbf{x} = (x_1, \dots, x_n)$ de taille n
- Échantillon $\Longrightarrow$ les x_i = réalisations de variables aléatoires X_i
- Échantillon i.i.d. $\Longrightarrow$ les X_i sont mutuellement indépendants

$$\Longrightarrow P(X_1 = x_1, \dots, X_n = x_n | \theta = \theta) = \prod_{i=1}^{n} P(X_i = x_i | \theta = \theta)$$

Vraisemblance d'un échantillon dans le cas discret

- $L(\mathbf{x}, \theta)$ = Vraisemblance de l'échantillon
- $L(\mathbf{x}, \theta)$ = proba d'obtenir cet échantillon sachant que $\theta = \theta$

$$L(\mathbf{x}, \theta) = P(x_1, \dots, x_n | \theta = \theta) = \prod_{i=1}^{n} P(x_i | \theta = \theta)$$

Vraisemblance d'un échantillon : le cas continu

- Paramètre à estimer : θ
- Échantillon $\mathbf{x} = (x_1, \dots, x_n)$ de taille n
- Échantillon i.i.d. $\Longrightarrow$ les X_i sont mutuellement indépendants
- p : fonction de densité

$$\Longrightarrow p(X_1 = x_1, \dots, X_n = x_n | \theta = \theta) = \prod_{i=1}^{n} p(X_i = x_i | \theta = \theta)$$

Vraisemblance d'un échantillon dans le cas continu

- $L(\mathbf{x}, \theta)$ = Vraisemblance de l'échantillon
- $L(\mathbf{x}, \theta) = p(x_1, \dots, x_n | \theta = \theta) = \prod_{i=1}^{n} p(x_i | \theta = \theta)$

Exemple de vraisemblance (1/2)

- pièce de monnaie : $P(\text{Pile}) = 0,75$ et $P(\text{Face}) = 0,25$
- jet de la pièce $\Longrightarrow$ expérience de Bernoulli
- paramètre θ = proba de Pile $= 0,75 = \theta$

- échantillon 1 :

P	P	F	F	P	P	F	P	P	P

$$\Longrightarrow L(\mathbf{x}, \theta) = \prod_{i=1}^{7} P(\text{Pile}|\theta) \times \prod_{i=1}^{3} P(\text{Face}|\theta)$$

$$= 0,75^7 \times 0,25^3 \approx 0,002086$$

- échantillon 2 :

F	F	P	P	F	F	P	F	F	F

$$\Longrightarrow L(\mathbf{x}, \theta) = \prod_{i=1}^{3} P(\text{Pile}|\theta) \times \prod_{i=1}^{7} P(\text{Face}|\theta)$$

$$= 0,75^3 \times 0,25^7 \approx 0,000026$$

Exemple de vraisemblance (2/2)

- pièce de monnaie : $P(\text{Pile}) = ???$ et $P(\text{Face}) = ???$
- paramètre θ = proba de Pile $= ???$
- échantillon :

P	P	F	F	P	P	F	P	P	P

$$\Longrightarrow L(\mathbf{x}, \theta) = \prod_{i=1}^{7} P(\text{Pile}|\theta) \times \prod_{i=1}^{3} P(\text{Face}|\theta)$$

- $\theta_1 = 0,75 \Longrightarrow L(x, \theta_1) = 0,75^7 \times 0,25^3 \approx 0,002086$
- $\theta_2 = 0,5 \Longrightarrow L(x, \theta_2) = 0,5^7 \times 0,5^3 \approx 0,000976$
- $\theta_3 = 0,25 \Longrightarrow L(x, \theta_3) = 0,25^7 \times 0,75^3 \approx 0,000026$

$\Longrightarrow \theta_1$ plus vraisemblable que θ_2 ou θ_3

Estimateur du maximum de vraisemblance

Estimateur du maximum de vraisemblance

- X : variable aléatoire sur la population
- X suit une loi de proba de paramètre θ inconnu
- Θ : ensemble des valeurs possibles pour θ
- $\mathbf{x}$: échantillon i.i.d.
- $T = f(X) =$ *estimateur du maximum de vraisemblance*

  défini par $\mathbf{x} \longmapsto t = f(\mathbf{x}) = \operatorname*{Argmax}_{\theta \in \Theta} L(\mathbf{x}, \theta)$

$\Longrightarrow t =$ valeur θ de θ pour laquelle la proba d'observer $\mathbf{x}$ était la plus grande

Calcul du maximum de vraisemblance

Problème : comment calculer le maximum de vraisemblance ?

- $\operatorname*{Argmax}_{\theta \in \Theta} L(\mathbf{x}, \theta) = \operatorname*{Argmax}_{\theta \in \Theta} P(x_1, \ldots, x_n|\theta) = \operatorname*{Argmax}_{\theta \in \Theta} \prod_{i=1}^{n} P(x_i|\theta)$
- Certaines conditions de concavité et de dérivabilité

  $\Longrightarrow \operatorname*{Argmax}_{\theta \in \Theta} L(\mathbf{x}, \theta)$ obtenu lorsque $\dfrac{\partial L(\mathbf{x}, \theta)}{\partial \theta} = 0$

- $\operatorname*{Argmax}_{\theta \in \Theta} L(\mathbf{x}, \theta) = \operatorname*{Argmax}_{\theta \in \Theta} \ln L(\mathbf{x}, \theta) = \operatorname*{Argmax}_{\theta \in \Theta} \sum_{i=1}^{n} \ln P(x_i|\theta)$

$$\operatorname*{Argmax}_{\theta \in \Theta} \ln L(\mathbf{x}, \theta) = \text{log vraisemblance}$$

$\Longrightarrow \operatorname*{Argmax}_{\theta \in \Theta} L(\mathbf{x}, \theta)$ obtenu lorsque $\sum_{i=1}^{n} \dfrac{\partial \ln P(x_i|\theta)}{\partial \theta} = 0$

Max de vraisemblance et loi normale (1/2)

- $X \sim \mathcal{N}(\mu, \sigma^2)$; on suppose $\sigma = 1$
- paramètre θ = espérance μ
- loi normale $\Longrightarrow$ vraisemblance :

$$L(\mathbf{x}, \theta) = \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}(x_i - \theta)^2 \right\} \right]$$

- $\dfrac{\partial L(\mathbf{x}, \theta)}{\partial \theta} = 0 \Longleftrightarrow \dfrac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} = 0$
- $\ln L(\mathbf{x}, \theta) = -\dfrac{n}{2} \ln 2\pi - \dfrac{1}{2} \displaystyle\sum_{i=1}^{n} (x_i - \theta)^2$
- $\dfrac{\partial L(\mathbf{x}, \theta)}{\partial \theta} = 0 \Longleftrightarrow \displaystyle\sum_{i=1}^{n} (x_i - \theta) = 0 \Longleftrightarrow \theta = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$

Estimateur du maximum de vraisemblance : $\overline{X}$

Max de vraisemblance et loi normale (2/2)

- $X \sim \mathcal{N}(\mu, \sigma^2)$
- paramètre $\theta = (\mu, \sigma^2)$
- Log vraisemblance :

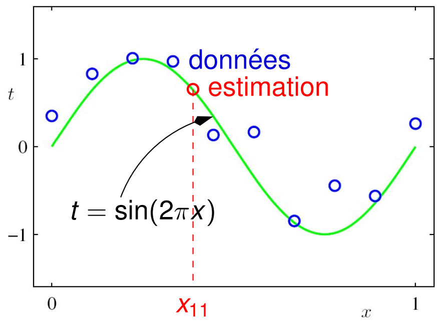
$$\ln L(\mathbf{x}, \theta) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

- Maximum de vraisemblance $\Longrightarrow \dfrac{\partial L(\mathbf{x}, \theta)}{\partial \mu} = 0$ et $\dfrac{\partial L(\mathbf{x}, \theta)}{\partial \sigma^2} = 0$
- $\begin{cases} \dfrac{\partial L(\mathbf{x}, \theta)}{\partial \mu} = \dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} (x_i - \mu) = 0 & \Longrightarrow \mu = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i = \overline{x} \\ \dfrac{\partial L(\mathbf{x}, \theta)}{\partial \sigma^2} = -\dfrac{n}{2} \dfrac{1}{\sigma^2} + \dfrac{1}{2\sigma^4} \displaystyle\sum_{i=1}^{n} (x_i - \mu)^2 = 0 & \Longrightarrow \sigma^2 = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} (x_i - \overline{x})^2 = s_n^2 \end{cases}$

Estimateurs du maximum de vraisemblance : $\overline{X}$ et S_n^2

⚠ estimateur de la variance biaisé : variance non corrigée

Retour sur le problème d'ajustement du cours 1 (1/6)



Observations
(x_1, t_1)
$\vdots$
(x_{10}, t_{10})

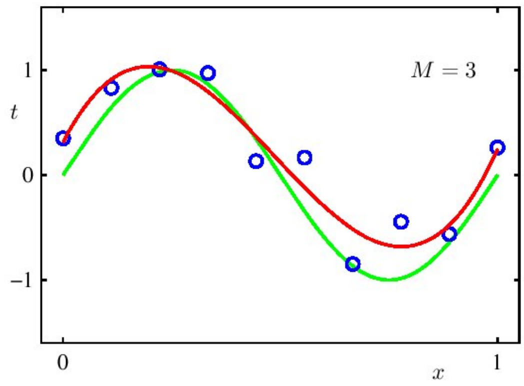
$\Longrightarrow$ courbe $\sin(2\pi x)$ $\Longrightarrow$ estimation de t_{11}

$\Longrightarrow$ reconnaissance de la courbe verte

Retour sur le problème d'ajustement du cours 1 (2/6)

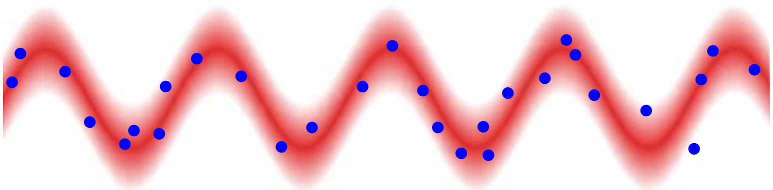
Idée : estimer la courbe verte par un polynôme :

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$



Retour sur le problème d'ajustement du cours 1 (3/6)

Idée : les ordonnées des points bleus sont distribuées selon une loi normale autour de $y(x, \mathbf{w})$:



$\Longrightarrow P(t|x, \mathbf{w}, \sigma^2) = \mathcal{N}(t|y(x, \mathbf{w}), \sigma^2)$

Problème : comment trouver $\mathbf{w}$ et σ^2 ?

$\Longrightarrow$ par maximum de vraisemblance

Retour sur le problème d'ajustement du cours 1 (4/6)

$$P(t|x, \mathbf{w}, \sigma^2) = \mathcal{N}(t|y(x, \mathbf{w}), \sigma^2)$$

- observations $\{(x_i, t_i), i = 1, \dots, n\}$
- $\mathbf{t} = \{t_1, \dots, t_n\}$; $\mathbf{x} = \{x_1, \dots, x_n\}$
- observations $\Longrightarrow$ échantillon i.i.d

$$\Longrightarrow P(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2) = \prod_{i=1}^{n} P(t_i|x_i, \mathbf{w}, \sigma^2)$$
$$= \prod_{i=1}^{n} \mathcal{N}(t_i|y(x_i, \mathbf{w}), \sigma^2)$$

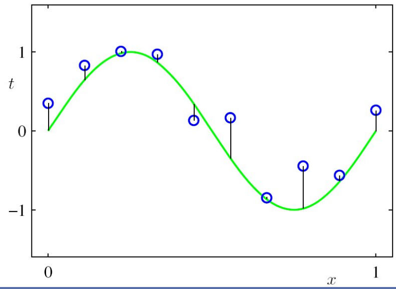
- Max de vraisemblance $\Longrightarrow$ calculer la log-vraisemblance :

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [y(x_i, \mathbf{w}) - t_i]^2 + \frac{n}{2} \ln \frac{1}{\sigma^2} - \frac{n}{2} \ln(2\pi)$$

Retour sur le problème d'ajustement du cours 1 (5/6)

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y(x_i, \mathbf{w}) - t_i]^2 + \frac{n}{2}\ln\frac{1}{\sigma^2} - \frac{n}{2}\ln(2\pi)$$

- Maximum de log-vraisemblance $\Longrightarrow$ trouver $\mathbf{w}_{ML}$ et σ^2_{ML} qui maximisent $\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2)$
- maximiser par rapport à $\mathbf{w}_{ML}$ $\Longleftrightarrow$ minimiser $\sum_{i=1}^{n}[y(x_i, \mathbf{w}) - t_i]^2$



$\Longrightarrow$ critère du cours 1.1

Retour sur le problème d'ajustement du cours 1 (6/6)

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y(x_i, \mathbf{w}) - t_i]^2 + \frac{n}{2}\ln\frac{1}{\sigma^2} - \frac{n}{2}\ln(2\pi)$$

- maximiser $\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2)$ par rapport à $\sigma^2 \Longrightarrow \dfrac{\partial \ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2)}{\partial \sigma^2} = 0$
- $\dfrac{\partial \ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2)}{\partial \sigma^2} = \dfrac{1}{2\sigma^4}\sum_{i=1}^{n}[y(x_i, \mathbf{w}) - t_i]^2 - \dfrac{n}{2\sigma^4}\sigma^2 = 0$

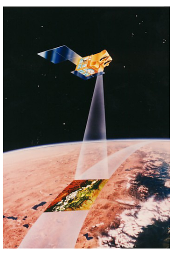
$$\Longrightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}[y(x_i, \mathbf{w}) - t_i]^2$$

$$\sigma^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}[y(x_i, \mathbf{w}_{ML}) - t_i]^2$$

Prévention des risques d'inondation (1/4)

- Plan de prévention des risques d'inondations (PPR-I) :

photos satellite SPOT5 $\Longrightarrow$ zones susceptibles d'être inondées



- 3 catégories de parcelles :
  1. inondables (*PI*)
  2. partiellement inondables (*PPI*)
  3. non inondables (*NI*)

Prévention des risques d'inondation (2/4)

- images en teintes de gris
- proba d'obtenir un niveau de gris n dépend du type de zone :

$$P(n|PI) = \mathcal{N}(\mu_1, \sigma_1^2) \qquad P(n|PPI) = \mathcal{N}(\mu_2, \sigma_2^2)$$
$$\mu_1 = 100 \quad \sigma_1 = 20 \qquad \mu_2 = 85 \quad \sigma_2 = 5$$

- nouvelle image envoyée par SPOT5 :

- zone Z : niveau de gris $= n = 80$

Problème : zone $Z = PI$ ou PPI ?

Prévention des risques d'inondation (3/4)

Problème : zone $Z = PI$ ou PPI ?

- 2 hypothèses :
  1. $\theta_1 =$ « Z est de type PI »
  2. $\theta_2 =$ « Z est de type PPI »
- Idée : calcul du max de vraisemblance d'obtenir la zone Z sous θ_1 ou sous θ_2
- $L(\mathbf{x}, \theta_1) = p(80|PI)$, avec p fct de densité de $P(n|PI) = \mathcal{N}(\mu_1, \sigma_1^2)$

Rappel : la fonction de densité de $\mathcal{N}(\mu, \sigma^2)$ est :

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

Prévention des risques d'inondation (4/4)

Problème : zone $Z = PI$ ou PPI ?

- $P(n|PI) = \mathcal{N}(\mu_1, \sigma_1^2) = \mathcal{N}(100, 20^2)$
- $L(\mathbf{x}, \theta_1) = p(80|PI)$

$$= \frac{1}{\sqrt{2\pi}\times 20} \exp\left\{ -\frac{1}{2} \left(\frac{80-100}{20} \right)^2 \right\}$$
$$= \frac{1}{20\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \right\}$$
$$\approx 0,0121$$

- $P(n|PPI) = \mathcal{N}(\mu_2, \sigma_2^2) = \mathcal{N}(85, 5^2)$
- $L(\mathbf{x}, \theta_2) = p(80|PPI) = \frac{1}{\sqrt{2\pi}\times 5} \exp\left\{ -\frac{1}{2} \left(\frac{80-85}{5} \right)^2 \right\} \approx 0,0484$

Max de vraisemblance $\Longrightarrow$ PPI plus probable

RFIDEC — cours 5: MAP et apprentissage non paramétrique

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours n°5

1. Maximum a posteriori
2. Estimation de densité

Max de vraisemblance et loi binomiale (1/2)

- pièce de monnaie
- $X \in \{0, 1\}$, $0 \Longleftrightarrow$ Face, $1 \Longleftrightarrow$ Pile
- $X \sim \mathcal{B}(1, p) \Longrightarrow P(X = x|p) = p^x(1-p)^{1-x}$
- n lancers de la pièce $\Longrightarrow$ observations $\mathbf{x} = \{x_1, \ldots, x_n\}$
- $P(\mathbf{x}|p) = \prod_{i=1}^{n} P(x_i|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

Problème : à partir de $\mathbf{x}$, peut-on raisonnablement déduire p ?

- maximum de vraisemblance :

$$\ln P(\mathbf{x}|p) = \sum_{i=1}^{n} [x_i \ln p + (1-x_i)\ln(1-p)]$$

$$\frac{\partial \ln P(\mathbf{x}|p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0 \Longrightarrow p_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Max de vraisemblance et loi binomiale (2/2)

$$p_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

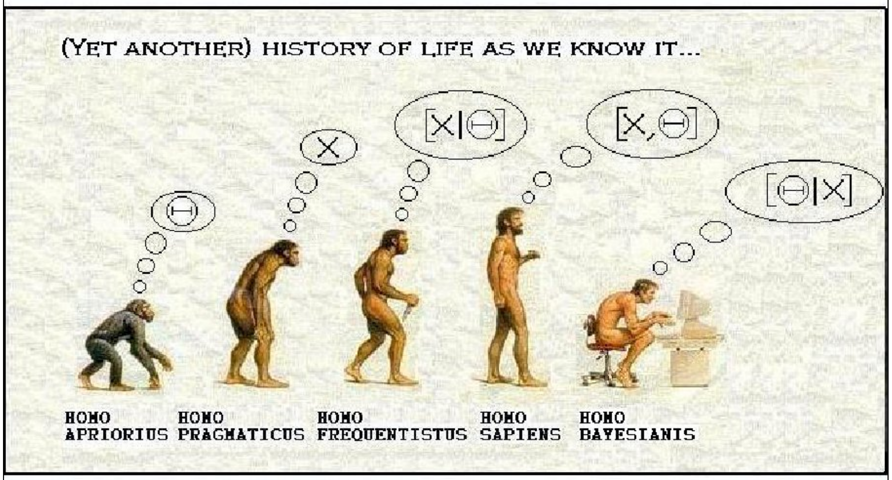
- 3 lancers $\Longrightarrow$ observations : {Pile,Pile,Pile}
- Maximum de vraisemblance $\Longrightarrow p_{ML} = 1$

$\Longrightarrow$ on considère que tout lancer de la pièce devrait tomber sur Pile

$\Longrightarrow$ résultat à l'encontre du bon sens

$\Longrightarrow$ autre estimateur : maximum a posteriori

En route vers le maximum a posteriori



Le modèle bayésien (1/4)

Maximum a posteriori $\Longrightarrow$ modèle bayésien

Modèle bayésien

événements : parties de $\mathcal{X} \times \Theta$, où :

- $\mathcal{X} =$ l'espace des observations (échantillons) $\mathbf{x}$ de taille n
- $\Theta =$ espace des paramètres θ
- famille des événements dotée d'une loi de proba Π

- *cas discret :* Π déterminée par les probas des événements élémentaires $\pi(\mathbf{x}, \theta)$
- *cas continu :* Π déterminée par la densité jointe $\pi(\mathbf{x}, \theta)$

⚠ Max de vraisemblance : $\pi(\mathbf{x}|\theta)$ au lieu de $\pi(\mathbf{x}, \theta) = \pi(\mathbf{x}|\theta)\pi(\theta)$

Le modèle bayésien (2/4)

Le cas discret :

- $\pi(\mathbf{x}, \theta) = \Pi(X = \mathbf{x}, \boldsymbol{\theta} = \theta)$, où $X, \boldsymbol{\theta}$ variables aléatoires
- $\pi(\mathbf{x}) = \Pi(X = \mathbf{x}) = \sum_{\theta \in \Theta} \Pi(X = \mathbf{x}, \boldsymbol{\theta} = \theta) = \sum_{\theta \in \Theta} \pi(\mathbf{x}, \theta)$
- $\pi(\theta) = \Pi(\boldsymbol{\theta} = \theta) = \sum_{\mathbf{x} \in \mathcal{X}} \Pi(X = \mathbf{x}, \boldsymbol{\theta} = \theta) = \sum_{\mathbf{x} \in \mathcal{X}} \pi(\mathbf{x}, \theta)$
- $\pi(\mathbf{x}|\theta) = \Pi(X = \mathbf{x}|\boldsymbol{\theta} = \theta) = \dfrac{\pi(\mathbf{x}, \theta)}{\pi(\theta)}$
- $\pi(\theta|\mathbf{x}) = \Pi(\boldsymbol{\theta} = \theta|X = \mathbf{x}) = \dfrac{\pi(\mathbf{x}, \theta)}{\pi(\mathbf{x})}$

Probabilités a priori et a posteriori

- $\pi(\theta) =$ probabilité a priori de θ
- $\pi(\theta|\mathbf{x}) =$ probabilité a posteriori de θ

Le modèle bayésien (3/4)

Le cas continu :

- $\pi(\mathbf{x}) = \Pi(X = \mathbf{x}) = \int_{\theta \in \Theta} \Pi(X = \mathbf{x}, \boldsymbol{\theta} = \theta) d\theta = \int_{\theta \in \Theta} \pi(\mathbf{x}, \theta) d\theta$
- $\pi(\theta) = \Pi(\boldsymbol{\theta} = \theta) = \int_{\mathbf{x} \in \mathcal{X}} \Pi(X = \mathbf{x}, \boldsymbol{\theta} = \theta) d\mathbf{x} = \int_{\mathbf{x} \in \mathcal{X}} \pi(\mathbf{x}, \theta) d\mathbf{x}$
- $\pi(\mathbf{x}|\theta) = \dfrac{\pi(\mathbf{x}, \theta)}{\pi(\theta)}$
- $\pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}, \theta)}{\pi(\mathbf{x})}$

Le modèle bayésien (4/4)

Probabilités a priori et a posteriori

- $\pi(\theta)$ = probabilité a priori de $\boldsymbol{\theta}$
  = idée que l'on se fait de $\boldsymbol{\theta}$ avant observation
- $\pi(\theta|\mathbf{x})$ = probabilité a posteriori de $\boldsymbol{\theta}$
  = idée que l'on se fait de $\boldsymbol{\theta}$ après observation

- Formule de Bayes : $\pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\pi(\mathbf{x})}$

- $$\begin{cases} \textit{cas discret :} \ \pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\sum_{\theta \in \Theta} \pi(\mathbf{x}, \theta)} = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\sum_{\theta \in \Theta} \pi(\mathbf{x}|\theta)\pi(\theta)} \\ \textit{cas continu :} \ \pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\int_{\theta \in \Theta} \pi(\mathbf{x}, \theta) d\theta} = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\int_{\theta \in \Theta} \pi(\mathbf{x}|\theta)\pi(\theta) d\theta} \end{cases}$$

- *Rappel :* $\pi(\mathbf{x}|\theta)$ = vraisemblance de l'échantillon $= L(\mathbf{x}, \theta)$

Maximum a posteriori

Maximum a posteriori (MAP)

T estimateur du maximum a posteriori de θ :
défini par $\mathbf{x} \longmapsto t = \underset{\theta \in \Theta}{\text{Argmax}}\, \pi(\theta|\mathbf{x})$

- échantillon i.i.d de n observations
- $X = (X_1, \ldots, X_n) \Longrightarrow \mathbf{x} = (x_1, \ldots, x_n)$ observation de X
- $\begin{cases} \textit{cas discret :}\ \pi(\theta|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta)\pi(\theta)}{\sum_{\theta \in \Theta} L(\mathbf{x},\theta)\pi(\theta)} \\ \textit{cas continu :}\ \pi(\theta|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta)\pi(\theta)}{\int_{\theta \in \Theta} L(\mathbf{x},\theta)\pi(\theta)d\theta} \end{cases}$
- échantillon i.i.d $\Longrightarrow \pi(\mathbf{x}|\theta) = L(\mathbf{x},\theta) = \begin{cases} \prod_{i=1}^{n} P(x_i|\theta) & \text{(discret)} \\ \prod_{i=1}^{n} p(x_i|\theta) & \text{(continu)} \end{cases}$

MAP : retour sur la pièce de monnaie (1/2)

- pièce de monnaie $\Longrightarrow X \in \{0, 1\}$

  $0 \Longleftrightarrow$ Face  $1 \Longleftrightarrow$ Pile 
- $X \sim \mathcal{B}(1, \theta) \Longrightarrow P(X = x|\theta) = \theta^x (1-\theta)^{1-x}$
- échantillon $\mathbf{x}$ de 3 lancers $\Longrightarrow$ {Pile,Pile,Pile} 
- Max de vraisemblance $\Longrightarrow \theta_{ML} = 1$

  $\Longrightarrow$ tous les lancers devraient tomber sur Pile

---

- Modèle bayésien : $\Theta = \{\theta_1, \theta_2\}$

  $\theta_1 =$ « biais en faveur de Pile », $\theta_2 =$ « biais en faveur de Face »
- Info a priori : $\pi(\theta_1) = \frac{2}{3}$, $\pi(\theta_2) = \frac{1}{3}$

Problème : quelle est la valeur du maximum a posteriori ?

MAP : retour sur la pièce de monnaie (2/2)

- Info a priori : $\pi(\theta_1) = \frac{2}{3}$, $\pi(\theta_2) = \frac{1}{3}$
- $L(\mathbf{x},\theta_1) = \pi(\mathbf{x}|\theta_1) = \prod_{i=1}^{3} P(x_i|\theta_1) = \frac{2}{3}^3 \times (1 - \frac{2}{3})^0 = \frac{2}{3}^3 \approx 0,296$
- $L(\mathbf{x},\theta_2) = \pi(\mathbf{x}|\theta_2) = \prod_{i=1}^{3} P(x_i|\theta_2) = \frac{1}{3}^3 \times (1 - \frac{1}{3})^0 = \frac{1}{3}^3 \approx 0,037$
- $\pi(\theta_1|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_1)\pi(\theta_1)}{\sum_{\theta \in \Theta} L(\mathbf{x},\theta)\pi(\theta)} = \dfrac{\frac{2}{3}^3 \times \frac{2}{3}}{\frac{2}{3}^3 \times \frac{2}{3} + \frac{1}{3}^3 \times \frac{1}{3}} \approx 0,941$
- $\pi(\theta_2|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_2)\pi(\theta_2)}{\sum_{\theta \in \Theta} L(\mathbf{x},\theta)\pi(\theta)} \approx 0,059$

Max a posteriori : $\theta = \theta_1 \Longrightarrow X \sim \mathcal{B}(1, \theta_1) = \mathcal{B}(1, 0.941)$

⚠ probabilité que la pièce tombe sur Face $\neq 0$

MAP et les lois conjuguées

calcul de la distribution a posteriori : $\pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\int_{\theta\in\Theta}\pi(\mathbf{x}|\theta)\pi(\theta)d\theta}$

$\Longrightarrow$ si $\pi(\mathbf{x}|\theta)\pi(\theta)$ complexe analytiquement alors calcul de l'intégrale compliqué

Lois conjuguées

- $\pi(\theta)$: loi a priori
- $\pi(\mathbf{x}|\theta)$: fonction de vraisemblance
- $\pi(\theta|\mathbf{x})$: distribution a posteriori
- $\pi(\theta)$ et $\pi(\mathbf{x}|\theta)$ sont conjuguées si $\pi(\theta|\mathbf{x})$ appartient à la même famille de lois que $\pi(\theta)$

Lois conjuguées : exemple de la pièce de monnaie

- pièce de monnaie $\Longrightarrow X \in \{0,1\}$: $0 \Longleftrightarrow$  $1 \Longleftrightarrow$ 
- $X \sim \mathcal{B}(1,\theta) \Longrightarrow$ vraisemblance d'un échantillon :

  $\pi(\mathbf{x}|\theta) = \theta^x(1-\theta)^{n-x}$, avec $x = \#(x_i = 1)$

  $\Longrightarrow$ loi binomiale

Distribution de probabilité Beta

Loi Beta : $\text{Beta}(\theta, a, b) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$

avec $\Gamma(x) = \displaystyle\int_0^{+\infty} t^{x-1}e^{-t}dt$

Espérance $= \dfrac{a}{a+b}$ Variance $= \dfrac{ab}{(a+b)^2(a+b+1)}$

$\Longrightarrow$ loi Beta et loi binomiales conjuguées

Lois conjuguées : loi binomiale et loi Beta

- loi a priori : $\pi(\theta) = \text{Beta}(\theta, a, b) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$
- fonction de vraisemblance : $\pi(\mathbf{x}|\theta) = \theta^x(1-\theta)^{n-x}$, avec $x = \#(x_i = 1)$
- loi a posteriori : $\pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\sum_{\theta\in\Theta}\pi(\mathbf{x}|\theta)\pi(\theta)} \propto \pi(\mathbf{x}|\theta)\pi(\theta)$
- loi a posteriori : $\pi(\theta|\mathbf{x}) \propto \theta^{x+a-1}(1-\theta)^{b+n-x-1}$

$$\Longrightarrow \pi(\theta|\mathbf{x}) \sim \text{Beta}(\theta, x+a, b+n-x)$$

Comparaison MAP – maximum de vraisemblance

- pièce de monnaie $\Longrightarrow X \in \{0, 1\}$: $0 \Longleftrightarrow$  $1 \Longleftrightarrow$ 
- Max de vraisemblance :

  $\pi(\mathbf{x}|\theta) = \theta^x(1-\theta)^{n-x} \Longrightarrow \text{Beta}(\theta, x+1, n-x+1)$
- Max a posteriori :

  $\pi(\theta|\mathbf{x}) \propto \theta^{x+a-1}(1-\theta)^{b+n-x-1} \Longrightarrow \text{Beta}(\theta, x+a, n-x+b)$

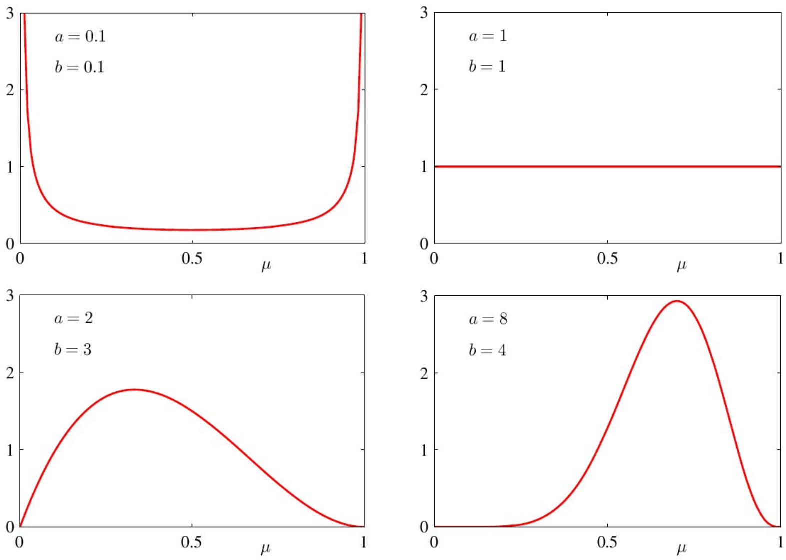
$\Longrightarrow$ Max de vraisemblance $\Longleftrightarrow$ Max a posteriori avec $a = 1$ et $b = 1$

Or $\text{Beta}(\theta, 1, 1) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} =$ constante

Max de vraisemblance $\Longleftrightarrow$ Max a posteriori avec a priori uniforme

⚠ $n \to +\infty \Longrightarrow$ max de vraisemblance $\approx$ max a posteriori
$\Longrightarrow$ l'a priori devient négligeable

La loi Beta



Loi normale et loi conjuguée

- fonction de vraisemblance = loi normale, σ^2 connue

  $\Longrightarrow$ loi a priori conjuguée : loi Γ

La loi Γ

- $X \sim \Gamma(x, k, \theta)$
- fonction de densité de la loi Γ :

$$f(x, k, \theta) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)} \quad \forall x, k, \theta > 0$$

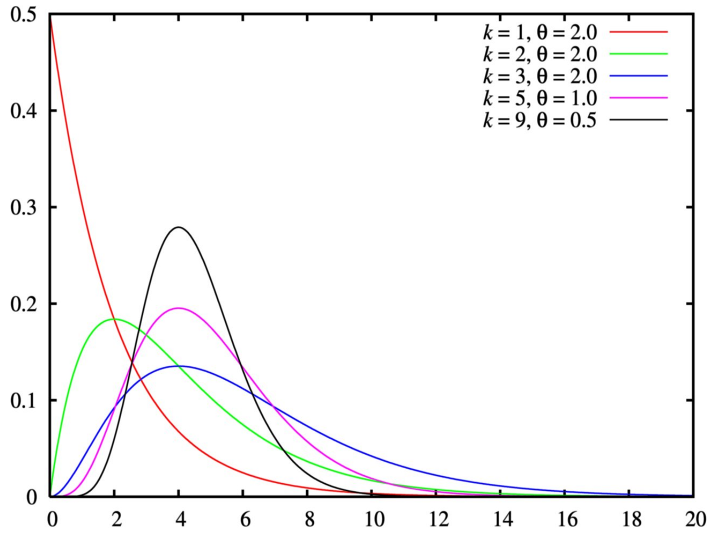
- $\Gamma(k) = \displaystyle\int_0^{+\infty} t^{k-1}e^{-t}dt$
- $E(X) = k\theta, \quad V(X) = k\theta^2$

⚠ Lorsque k entier : $\Gamma(x, k, \theta) =$ loi de k variables indépendantes suivant une loi exponentielle d'espérance θ

- Familles de lois conjuguées :

  `http ://en.wikipedia.org/wiki/Conjugate_prior`

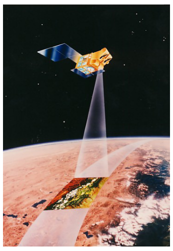
Loi Gamma



Prévention des risques d'inondation (1/3)

- Plan de prévention des risques d'inondations (PPR-I) :

photos satellite SPOT5 $\Longrightarrow$ zones susceptibles d'être inondées



- 3 catégories de parcelles :
  1. inondables (*PI*)
  2. partiellement inondables (*PPI*)
  3. non inondables (*NI*)

Prévention des risques d'inondation (2/3)

- images en teintes de gris
- proba d'obtenir un niveau de gris n dépend du type de zone :

$$P(n|PI) = \mathcal{N}(100, 20^2) \qquad P(n|PPI) = \mathcal{N}(85, 5^2)$$

- nouvelle image envoyée par SPOT5 :

- zone Z : niveau de gris $= n = 80$
- Connaissance a priori : 60% de *PI*, 10% de *PPI*, 30% de *NI*

Problème : zone $Z = PI$ ou PPI ?

Prévention des risques d'inondation (3/3)

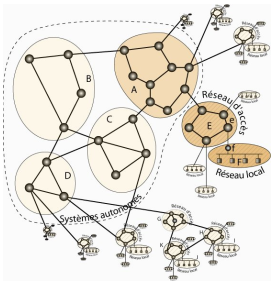
Problème : zone $Z = PI$ ou PPI ?

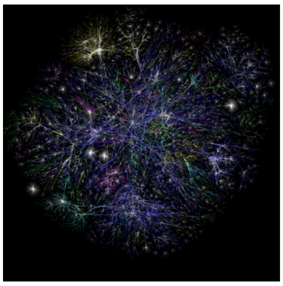
- 2 hypothèses :
  1. $\theta_1 =$ « Z est de type PI »
  2. $\theta_2 =$ « Z est de type PPI »
- Idée : calcul du MAP d'obtenir la zone Z sous θ_1 ou sous θ_2
- $\pi(\theta_1|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_1)\pi(\theta_1)}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)}$ $\qquad \pi(\theta_2|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_2)\pi(\theta_2)}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)}$
- Rappel cours 4 : $L(\mathbf{x},\theta_1) \approx 0,0121 \qquad L(\mathbf{x},\theta_2) \approx 0,0484$
- a priori : $\pi(\theta_1) = 0,6 \qquad \pi(\theta_2) = 0,1$
- $\pi(\theta_1|\mathbf{x}) = \dfrac{0,0121 \times 0,6}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)}$ $\qquad \pi(\theta_2|\mathbf{x}) = \dfrac{0,0484 \times 0,1}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)}$

MAP $\Longrightarrow$ parcelle inondable (PI)

Analyse d'un trafic réseau (1/4)

- Réseau informatique : transfert de paquets





- Problème : analyse des paquets perdus sur un sous-réseau
- X : variable aléatoire « nombre de paquets envoyés jusqu'à bonne réception »
- X loi géométrique : $P(X = n) = (1-p)^{n-1}p$

  p : probabilité qu'un paquet soit correctement transmis

Analyse d'un trafic réseau (2/4)

- observation de 7 réalisations de X :

2	3	8	3	4	7	8

Estimation de p ?

1. estimation par max de vraisemblance
2. estimation par MAP

Analyse d'un trafic réseau (3/4)

1 estimation par max de vraisemblance

2	3	8	3	4	7	8

- vraisemblance : $L(\mathbf{x}, \theta) = \prod_{i=1}^{7} P(x_i|\theta)$

  $\theta =$ estimation de p
- observations $\Longrightarrow L(\mathbf{x}, \theta) = (1-\theta)^{28}\theta^7$

$$\Longrightarrow \ln L(\mathbf{x}, \theta) = 28 \ln(1-\theta) + 7 \ln \theta$$

$$\Longrightarrow \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} = \frac{-28}{1-\theta} + \frac{7}{\theta} = \frac{7-35\theta}{p(1-\theta)}$$

$$\Longrightarrow \text{maximum de vraisemblance} = \theta = 0,2$$

Analyse d'un trafic réseau (4/4)

2 estimation par max de vraisemblance

2	3	8	3	4	7	8

- A priori : $\pi(\theta) = \text{Beta}(\theta, 2, 15) = \frac{\Gamma(17)}{\Gamma(2)\Gamma(15)}\theta^1(1-\theta)^{14}$
- $\text{Argmax}_\theta \pi(\theta|\mathbf{x}) = \text{Argmax}_\theta L(\mathbf{x}, \theta)\pi(\theta)$

$$= \text{Argmax}_\theta [(1-\theta)^{28}\theta^7] \times [(1-\theta)^{14}\theta]$$

$$= \text{Argmax}_\theta (1-\theta)^{42}\theta^8$$

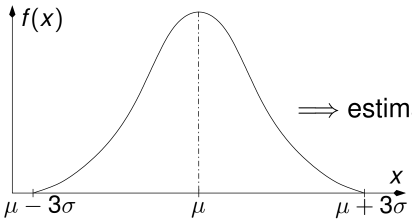
$$= \text{Argmax}_\theta\ 42 \ln(1-\theta) + 8 \ln \theta$$

$$\Longrightarrow \theta_{MAP} = 0,16$$

2 Estimation de densité

En route vers le non paramétrique

- jusqu'à maintenant : estimations des paramètres des densités :
  - max de vraisemblance
  - max a posteriori



$\Longrightarrow$ estimation de μ, σ^2

$\Longrightarrow$ *statistique paramétrique*

$\Longrightarrow$ connaissance de la famille de lois a priori

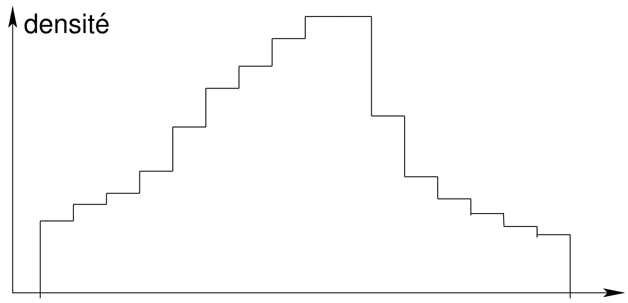
Que faire si aucune loi a priori évidente ?

$\Longrightarrow$ statistique non paramétrique

Estimation de densité : idée naïve

- X : variable aléatoire
- $\{x_1, \dots, x_n\}$: les classes de X
- échantillon de taille $n \Longrightarrow n_i$ individus dans la classe x_i

Densité $\approx$ histogramme des valeurs observées

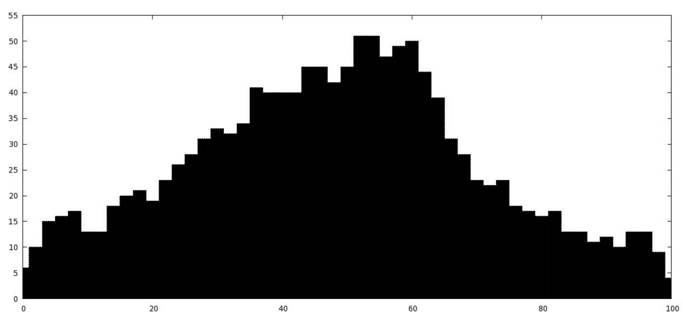


Problème : fonction de densité très irrégulière

Estimation de densité : la fenêtre mobile (1/2)

Idée de la fenêtre mobile

- X : variable aléatoire
- $\forall x$ valeur de X, $I_x = [x - \frac{h}{2}, x + \frac{h}{2}[$ = classe de longueur h
- densité estimée : $\hat{f}(x) = \frac{1}{n} \#x_i \ : x - \frac{h}{2} \leq x_i < x + \frac{h}{2}$



Problème : fonction de densité encore trop irrégulière

Estimation de densité : la fenêtre mobile (2/2)

densité estimée : $\hat{f}(x) = \frac{1}{n}\,\#x_i \;: x - \frac{h}{2} \leq x_i < x + \frac{h}{2}$

$$\Longrightarrow \hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

où K = fonction indicatrice de l'intervalle $\left[-\frac{1}{2}, \frac{1}{2}\right[$:

$$K(u) = \begin{cases} 0 & \text{si } u \geq \frac{1}{2} \text{ ou } u < -\frac{1}{2} \\ 1 & \text{si } -\frac{1}{2} \leq u < \frac{1}{2} \end{cases}$$

Estimation de densité par la méthode du noyau

- X : variable aléatoire
- $\forall x$ valeur de X, $I_x = [x - \frac{h}{2}, x + \frac{h}{2}[$ = classe de longueur h
- densité estimée : $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$

  avec K = fonction continue = *noyau*

Estimation de densité par la méthode du noyau

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

- noyaux fréquemment utilisés :
  - noyau gaussien : $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$
  - noyau d'Epanechnikov : $K(u) = \frac{3}{4\sqrt{5}}\left(1 - \frac{u^2}{5}\right)$ pour $|u| < \sqrt{5}$
- constante h = constante de lissage

  $$\begin{cases} h \text{ petit} & \Longrightarrow \hat{f} \text{ très irrégulière} \\ h \text{ grand} & \Longrightarrow \hat{f} \text{ très (trop) lisse} \end{cases}$$

- h souvent défini empiriquement