

II. Bases en probabilités

Mémo : l'échantillonnage rend la réponse statistique aléatoire !

Préambule

1. Théorie des probabilités
2. Dénombrement
3. Variables aléatoires
4. Lois de probabilités usuelles

Préambule

Expérience à résultat **aléatoire**

réalise des **événements**

simples ou **composés**

qui ont une **probabilité**

à compter dans certains cas :
dénombrement

une valeur numérique
leur est donnée par une :
variable aléatoire

d'où des fonctions donnant les
probabilités d'ensembles de
nombres : **lois de probabilité**

II.1.Théorie des probabilités

- a) Expérience et événements aléatoires
- b) Notions de probabilités

II.1.a Expérience et événements aléatoires

Action { résultats } = E connu

produit 1 résultat de E imprévisible

= expérience aléatoire

Descriptions (de ce qui arrive)

Non ambiguë, 1 résultat

= événement simple

Vague, plusieurs résultats

= événement composé

II.1.a Expérience et événements aléatoires (suite)

{ événements } = $\mathcal{E}$ = { parties de E }

Composition des événements par $\cup$, $\cap$, -

Evénements spéciaux

impossible = $\emptyset$

certain = E

complémentaire de A = $\overline{A}$

II.1.b Notions de probabilités

- α) Définition, mesure de probabilité
- β) Axiomes de la théorie
- γ) Théorèmes importants
- δ) Affectation d'une valeur de probabilité

II.1.b.α Définition, mesure de probabilité

Quantification de la certitude = probabilité

$$P : \mathcal{E} \rightarrow [0 ; 1]$$

$$A \rightarrow P(A)$$

= mesure de proba

II.1.b.β Axiomes de la théorie

$$A1) \quad P(E)=1$$

$$A2) \quad \forall A \in \mathcal{E} \quad P(A) \geq 0$$

$$A3) \quad \forall A, B \in \mathcal{E} \quad A \cap B = \emptyset \quad P(A \cup B) = P(A) + P(B)$$

$$P(A) + P(\overline{A}) = 1$$

$$P(\emptyset) = 0$$

$$\text{Si } A \subseteq B \text{ alors } P(A) \leq P(B)$$

$$0 \leq P(A) \leq 1$$

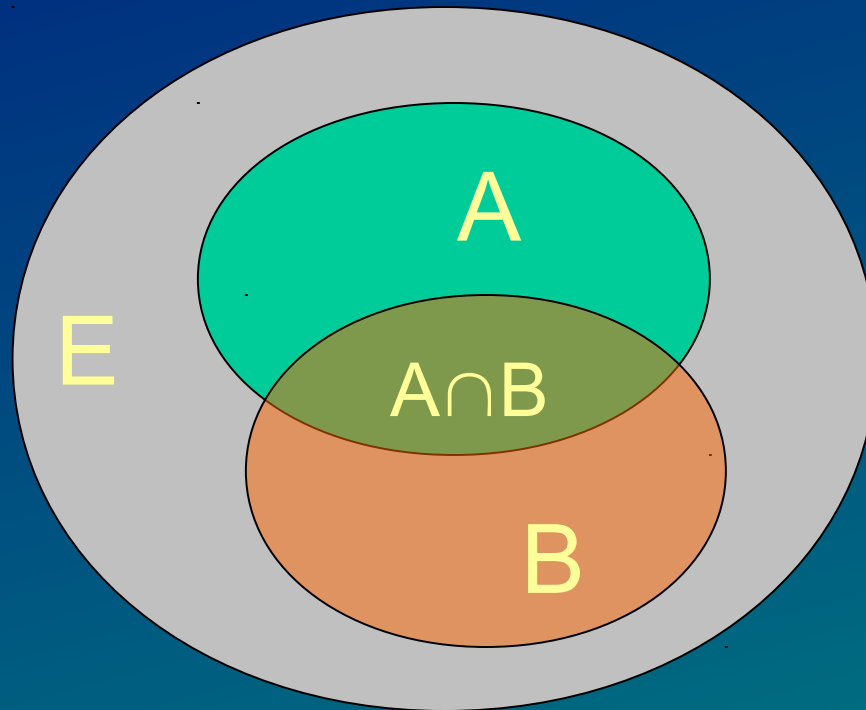
Conséquences :

II.1.b.γ Théorèmes importants

- i* Probabilités totales
- ii* Probabilités composées
- iii* Événements indépendants
- iii* Formule de Bayes

II.1.b.γ.i Probabilités totales

$$\forall A, B \in \mathcal{E} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



II.1.b.γ.ii Probabilités composées

Probabilité conditionnelle

(proba de B sachant A)

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

D'où :

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

II.1.b.γ.iii Événements indépendants

A indépendant de B :

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

Dans ce cas :

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

II.1.b.γ.iiii Formule de Bayes

K causes C_i Faits (conséquents) F_j :

$$P(C_i|F_j) = \frac{P(F_j|C_i)P(C_i)}{\sum_{k=1}^K P(F_j|C_k)P(C_k)}$$

II.1.b.δ Affectation d'une valeur de probabilité

Equiprobabilité :

N résultats (R_k)

$$P(R_k) = 1 / N \quad \forall k$$

P.S. Laplace

Fréquence limite :

N_x expériences

N_k fois R_k

$$F(R_k) = N_k / N_x$$

$$P(R_k) = \lim F(R_k)$$

quand $N_x \rightarrow \infty$

J.Neyman

Applications du II.1

- 1) Jet d'une pièce
- 2) Jet d'un dé
- 3) Tirage d'une carte
- 4) Observation obésité-hypertension
- 5) Test diagnostic, observation maladie-test
- 6) Association de tests
- 7) Compromis sensibilité-spécificité

1) Jet d'une pièce

$$E = \{ \textit{pile} , \textit{face} \}$$

$$\mathcal{E} = \{ \emptyset , \{ \textit{pile} \} , \{ \textit{face} \} , E \}$$

Equiprobabilité (pièce équilibrée) :

$$P(\textit{pile}) = P(\textit{face}) = 0,5$$

2) Jet d'un dé

$$E = \{ F1 , F2 , F3 , F4 , F5 , F6 \}$$

$$\mathcal{E} = \{ \emptyset , \{ F1 \} , \{ F2 \} , \dots , \{ F6 \} , \\ \{ F1 , F2 \} , \{ F1 , F3 \} , \dots , \{ F5 , F6 \} , \\ \{ F1 , F2 , F3 \} , \dots , \\ \dots , \{ F2 , F3 , F4 , F5 , F6 \} , \quad E \}$$

Equiprobabilité (dé équilibré) :

$$P(F_k) = 1 / 6 \quad \forall k$$

2) Jet d'un dé (suite)

Événements composés :

$$A = \text{« n° de face } \geq 3 \text{ »} = \{ F3, F4, F5, F6 \}$$

$$P(A) = 4 / 6 = 0,667$$

$$B = \text{« n° de face pair »} = \{ F2, F4, F6 \}$$

$$P(B) = 3 / 6 = 0,5$$

$$A \cap B = \{ F4, F6 \}$$

$$P(A \cap B) = 2 / 6 = 0,333$$

2) Jet d'un dé (suite)

Probabilités totales :

$$P(A \cup B) = 4/6 + 3/6 - 2/6 = 5/6 = 0,833$$

De fait :

$$A \cup B = \{ F2, F3, F4, F5, F6 \}$$

3) Tirage d'une carte

$$E = \{ \text{Roi } \heartsuit, \text{ Dame } \heartsuit, \text{ Valet } \heartsuit, \dots, 7 \clubsuit, \text{ As } \clubsuit \}$$

$$\mathcal{E} = \{ \emptyset, \{ \text{Roi } \heartsuit \}, \dots \text{ bref ! } \dots, E \}$$

Equiprobabilité :

$$P(\text{ Carte }) = 1 / 32 \quad \forall \text{ Carte}$$

$$P(\text{ As } \cup \heartsuit) = P(\text{ As ou } \heartsuit)$$

$$= P(\text{ As }) + P(\heartsuit) - P(\text{ As } \cap \heartsuit)$$

$$= 4 / 32 + 8 / 32 - 1 / 32 = 0,344$$

3) Tirage d'une carte (suite)

$$P(\text{As} \mid \text{rouge})$$

$$= P(\text{As} \cap \text{rouge}) / P(\text{rouge})$$

$$= P(\{ \text{As} \heartsuit, \text{As} \spadesuit \}) / P(\text{rouge})$$

$$= (2 / 32) / (16 / 32)$$

$$= 0,125$$

4) Observation obésité-hypertension

$$E = \{ \text{Ob Hy} , \text{Ob } \underline{\text{Hy}} , \underline{\text{Ob}} \text{ Hy} , \underline{\text{Ob}} \underline{\text{Hy}} \}$$

N = 10_000 observations :

	Hy	<u>Hy</u>	
Ob	1_000	2_000	3_000
<u>Ob</u>	500	6_500	7_000
	1_500	8_500	10_000

4) Observation obésité-hypertension (suite)

Estimation des probabilités :

$$P(Ob Hy) = 1_000 / 10_000 = 0,1$$

$$P(Ob \underline{Hy}) = 2_000 / 10_000 = 0,2$$

$$P(\underline{Ob} Hy) = 500 / 10_000 = 0,05$$

$$P(\underline{Ob} \underline{Hy}) = 6_500 / 10_000 = 0,65$$

$$\begin{aligned} P(Ob \cup Hy) &= P(Ob) + P(Hy) - P(Ob \cap Hy) \\ &= [P(Ob Hy) + P(Ob \underline{Hy})] \\ &\quad + [P(\underline{Ob} Hy) + P(\underline{Ob} \underline{Hy})] \\ &\quad - P(Ob Hy) \\ &= 0,35 \end{aligned}$$

5) Test diagnostic, observation maladie-test

$$E = \{ MT, M\bar{T}, \underline{M}T, \underline{M}\bar{T} \}$$

N Observations :

	M	<u>M</u>	
T	VP	FP	VP + FP
<u>T</u>	FN	VN	FN + VN
	VP + FN	FP + VN	N

5) Test diagnostic, observation maladie-test (suite)

Prévalence :

$$P(\underline{M}) = (\text{VP} + \text{FN}) / N$$

Sensibilité :

$$\text{Se} = P(\underline{T} \mid \underline{M}) = \text{VP} / (\text{VP} + \text{FN})$$

Spécificité :

$$\text{Sp} = P(\underline{\bar{T}} \mid \underline{\bar{M}}) = \text{VN} / (\text{VN} + \text{FP})$$

5) Test diagnostic, observation maladie-test (suite)

Valeur prédictive positive :

$$VPP = P(M | T)$$

$$= \frac{Se P(M)}{Se P(M) + (1 - Sp) P(\underline{M})}$$

Valeur prédictive négative :

$$VPN = P(\underline{M} | \underline{T})$$

$$= \frac{Sp P(\underline{M})}{Sp P(\underline{M}) + (1 - Se) P(M)}$$

5) Test diagnostic, observation maladie-test (suite)

Efficienne du test :

$$\begin{aligned} Ef &= P(T \text{ et } M) + P(\underline{T} \text{ et } \underline{M}) \\ &= Se P(M) + Sp P(\underline{M}) \end{aligned}$$

Test informatif :

$$\text{si } P(M | T) > P(M)$$

$$\text{ou } P(\underline{M} | \underline{T}) > P(\underline{M})$$

donc si :

$$Y = Se + Sp - 1 \in [0; 1]$$

5) Test diagnostic, observation maladie-test (suite)

par exemple :

prévalence :

$$P(M) = 0,001 \text{ (1‰)}$$

sensibilité :

$$Se = 0,934$$

taux de faux positifs :

$$\begin{aligned} TFP &= P(T | \underline{M}) = 1 - Sp \\ &= 0,002 \end{aligned}$$

alors :

$$VPP = 0,318 \text{ (médiocre, dû à } P(M))$$

6) Association de tests

Tests A et B associés en série T :

$$T = A \text{ et } B$$

$$\bar{T} = (\bar{A} \text{ ou } \bar{B})$$

$$= (\bar{A} \text{ et } B) \text{ ou } (A \text{ et } \bar{B}) \text{ ou } (\bar{A} \text{ et } \bar{B})$$

Se diminuée

$$Se = P(A | M) P(B | M)$$

$$= Se_A Se_B < Se_A \text{ et } < Se_B$$

Sp augmentée

$$Sp = P(\bar{A} | \bar{M}) P(\bar{B} | \bar{M}) + P(A | \bar{M}) P(\bar{B} | \bar{M}) + P(\bar{A} | \bar{M}) P(B | \bar{M})$$

$$= Sp_A (1 - Sp_B) + Sp_B$$

$$= Sp_B (1 - Sp_A) + Sp_A > Sp_A \text{ et } > Sp_B$$

6) Association de tests (suite)

Tests A et B associés en parallèle T :

$$\begin{aligned} T &= (A \text{ ou } B) \quad \underline{T} = (\underline{A} \text{ et } \underline{B}) \\ &= (A \text{ et } \underline{B}) \text{ ou } (\underline{A} \text{ et } B) \text{ ou } (A \text{ et } B) \end{aligned}$$

Se augmentée

$$\begin{aligned} Se &= P(A | M) P(\underline{B} | M) + P(\underline{A} | M) P(B | M) + P(A | M) P(B | M) \\ &= Se_A (1 - Se_B) + Se_B \\ &= Se_B (1 - Se_A) + Se_A > Se_A \text{ et } > Se_B \end{aligned}$$

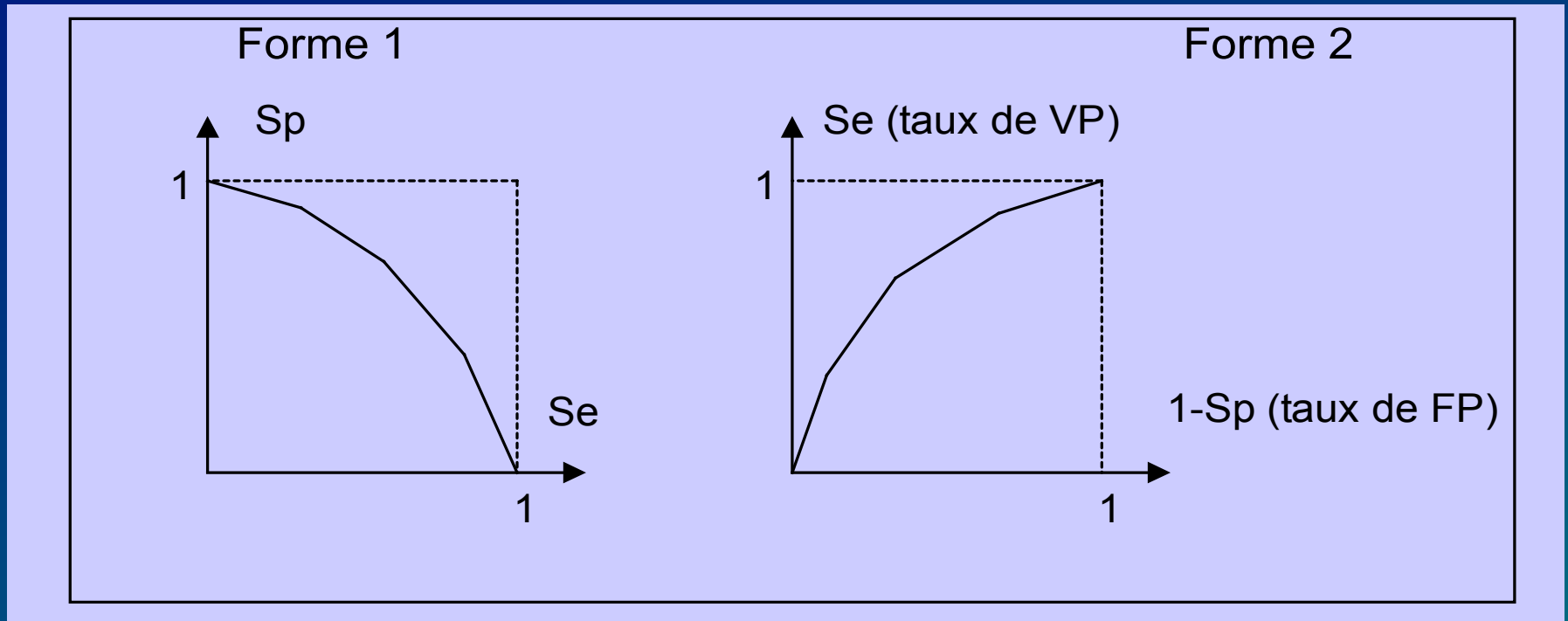
Sp diminuée

$$\begin{aligned} Sp &= P(\underline{A} | \underline{M}) P(\underline{B} | \underline{M}) \\ &= Sp_A Sp_B < Sp_A \text{ et } < Sp_B \end{aligned}$$

7) Compromis Se Sp

Se augmenté $\rightarrow$ diminution de Sp

Relation de la forme :



Forme 2 = courbe ROC

7) Compromis Se Sp (suite)

Compromis maladie grave

Se grande (max de VP, peu de FN) :

curable, FP non traumatisant.

Sp grande (peu de FP, max de VN)

incurable, FN sans incidence

VPP grande (max de M au test T)

traitement réservé aux malades

Ef grande

curable, FP et FN lourds de conséquences

II.2 Dénombrement

- a) Utilité
- b) Règle de base
- c) Permutations
- d) Arrangements
- e) Combinaisons

II.2.a Utilité

N résultats équiprobables

N_A résultats réalisant A

$$P(A) = N_A / N$$

Comptage direct impraticable

→ Règles de dénombrement

II.2.b Règle de base

succession de N épreuves e_i

e_i peut produire n_i résultats

Nombre de successions différentes de N résultats :

$$S = \prod_{i=1}^N n_i$$

II.2.c Permutations

N objets

Mis en ordre (permutation)

Nombre de permutations différentes de N objets :

$$N! = \prod_{i=1}^N i$$

Règle de base avec **e_i** = prise d'un objet dans les N- (i-1)
restants

II.2.c Permutations (suite)

N objets répartis en **K** familles de n_k objets (k de 1 à K).

Nombre de permutations différentes de **N** objets

en comptant comme une les permutations d'objets identiques :

$$\frac{N!}{\prod_{k=1}^K n_k!}$$

II.2.d Arrangements

N objets

En choisir **r** , les ranger (permutation)

Nombre de permutations différentes de **r** objets pris sur **N** (arrangements):

$$A_N^r = \frac{N!}{(N - r)!}$$

II.2.e Combinaisons

N objets

En choisir **r** , sans ordre

Nombre de paquets différents de **r** objets pris sur **N**
(combinaisons):

$$C_N^r = \frac{N!}{r! (N - r)!}$$

II.2.e Combinaisons (suite)

N objets

En choisir **r** , sans ordre , en remettant à chaque fois celui pris.

Nombre de paquets différents de **r** objets pris avec remise sur **N**
(combinaisons avec remise):

$$C_{N+r-1}^r = \frac{(N + r - 1)!}{r! (N - 1)!}$$

Applications du II.2

- 1) Système de codes
- 2) Anagrammes
- 3) Traitement combiné
- 4) Echantillonnage

1) Système de codes

Combien de codes de 5 lettres ?

Règle de base avec 5 épreuves « choix d'une lettre sur 26 »

$$\text{Nb} = 26^5 = 11_881_376$$

Avec des lettres toutes différentes ?

Choix de 5 lettres prises dans les 26 et ordonnées :

$$A_{26}^5 = \frac{26!}{(26 - 5)!}$$

$$= 26 \times 25 \times 24 \times 23 \times 22 = 7_893_600$$

2) Anagrammes

Combien d'anagrammes avec « MALLEOLE » ?

Permutations de $N = 8$ lettres avec répétitions.

$K = 5$ familles :

$$\text{« M »} \quad n_1 = 1$$

$$\text{« A »} \quad n_2 = 1$$

$$\text{« L »} \quad n_3 = 3$$

$$\text{« E »} \quad n_4 = 2$$

$$\text{« O »} \quad n_5 = 1$$

$$Nb = 8! / (3! 2!)$$

$$= 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 / 2 = 3_360$$

3) Traitement combiné

Avec 4 antibiotiques, combien d'associations de 3 produits ?

Choix de $r = 3$ produits sur $N = 4$ sans ordre.

$$C_4^3 = \frac{4!}{3! (4 - 3)!} = 4$$

4) Echantillonnage

Combien de groupes de 100 personnes dans un ensemble de 20_000 ?

$$C_{20_000}^{100} = \frac{20_000!}{100!(20_000 - 100)!} = \frac{\prod_{i=0}^{99} (20_000 - i)}{100!}$$

Gigantesque $\approx 10^{272}$ (!!)

II.3 Variables aléatoires

- a) Définition
- b) Loi de probabilité
- c) V.A. discrète
- d) V.A. continue
- e) Caractéristiques
- f) Inégalité de Bienaymé-Tchebichev

II.3.a Définition

Association d'une valeur numérique réelle à un événement simple

= Variable aléatoire X

$$X : E \rightarrow \mathbb{R}$$

$$e \mapsto r$$

D'où aussi (correspondance de parties) :

$$\mathcal{X} : \mathcal{E} \rightarrow \mathcal{P}(\mathbb{R})$$

$$A \mapsto \{ r \mid X^{-1}(r) \in A \}$$

II.3.a Définition (suite)

Exemple jeu de dé : « si face F_k gain de 10 k euros »

Variable aléatoire X :

$$X : \{ F_1, F_2, \dots, F_6 \} \rightarrow \{ 10, 20, \dots, 60 \}$$

$$F_k \mapsto 10k$$

Correspondance de parties $\mathcal{X}$:

$$\mathcal{X}(\{ F_2, F_5 \}) = \mathcal{X}(A) = \{ 20, 50 \}$$

Car : $X^{-1}(20) = F_2 \in A$ et aussi : $X^{-1}(50) = F_5 \in A$

II.3.b Loi de probabilité

Une mesure de probabilité P définie sur $\mathcal{E}$ transposée sur $\mathcal{P}(\mathbb{R})$ à l'aide de la V.A. X

= Loi de probabilité L de X

$$\mathcal{E} \xrightarrow{X} \mathcal{P}(\mathbb{R})$$

$$P \searrow \swarrow L$$

$$[0; 1]$$

$$A \longrightarrow X(A)$$

$$\searrow \swarrow$$

$$P(A) = L(X(A))$$

II.3.b Loi de probabilité (suite)

Exemple jeu de dé : « si face F_k gain de 10 k euros »

$$P(F_1) = 1 / 6$$

$$L(10) = 1 / 6$$

$$P(\{F_2, F_5\}) = 2 / 6$$

$$L(\{20, 50\}) = 2 / 6$$

Etc...

II.3.c V.A. discrète

E constitué d'événements discrets

La loi de la V.A. X est définie par

$$\{ (x_i ; p_i) \}$$

Avec :

$$0 < p_i < 1$$

$$\sum_i p_i = 1$$

II.3.c V.A. discrète (suite)

Exemple jeu de dé : « si face F_k gain de 10 k euros »

Donner les $(x_i ; p_i)$:

$$\{ (10, 1/6), (20, 1/6), \dots, (60, 1/6) \}$$

II.3.c V.A. discrète (suite)

Fonction de répartition (loi cumulative)

Définie par :

$$\{ (x_i ; F(x_i)) \}$$

Avec :

$$F(x_i) = \sum_{j \leq i} p_j$$

II.3.c V.A. discrète (suite)

Exemple jeu de dé : « si face F_k gain de 10 k euros »

Donner les $(x_i ; F(x_i))$:

$$\{ (10, 1/6), (20, 2/6), (30, 1/2), \dots, (60, 1) \}$$

II.3.d V.A. continue

Si E est ensemble continu d'événements,
la loi de la V.A. X est définie par la densité

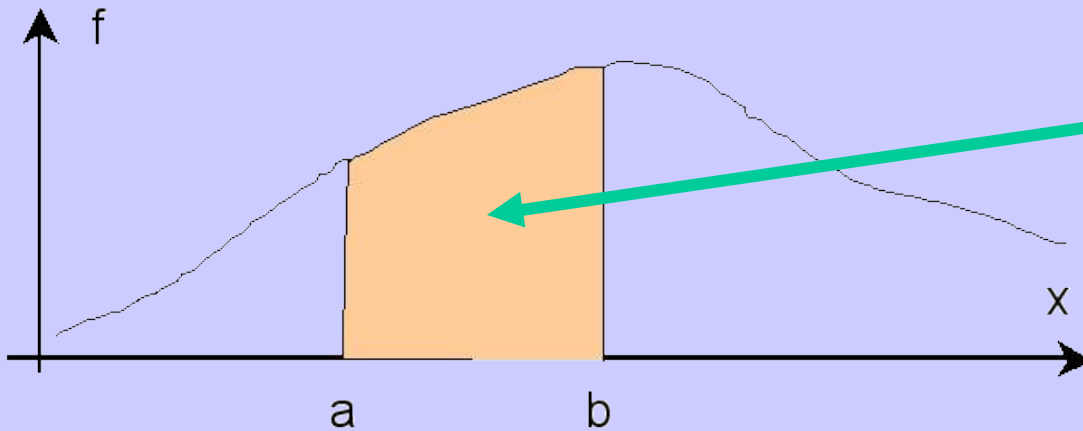
$$f(x)$$

Telle que :

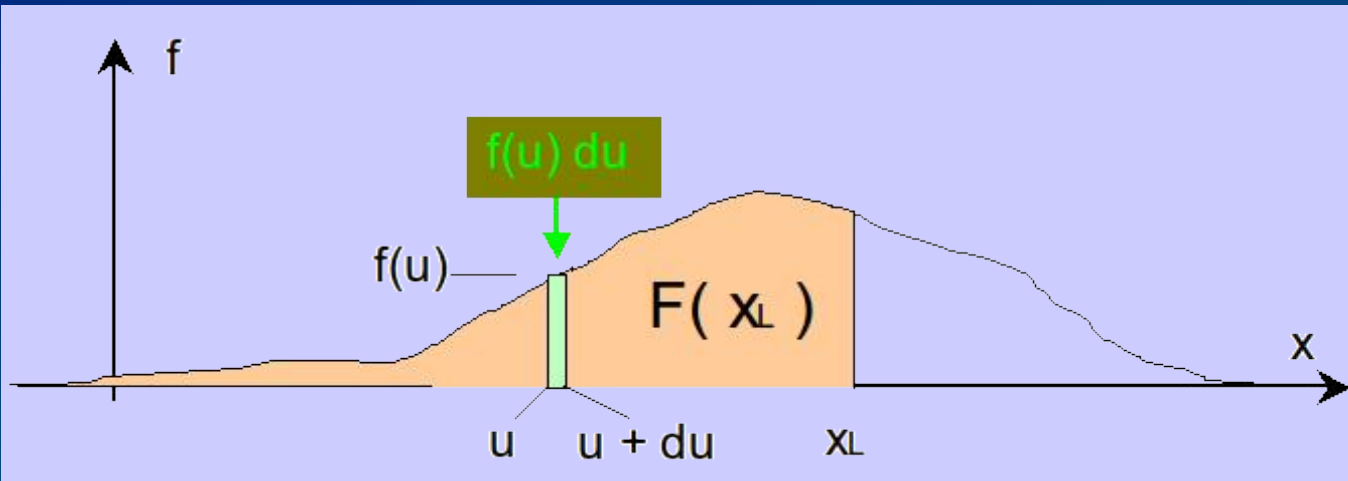
$$P(x < X < x + dx) = f(x) dx$$

Obligé de faire cela car maintenant $P(x) = 0$

II.3.d Note



$$\int_a^b f(x) dx$$



$$F(x_L) = \int_{-\infty}^{x_L} f(u) du$$

II.3.d V.A. continue (suite)

On a :

$$P(X \in [a, b]) = \int_a^b f(x) dx$$

$$P(X \in \mathbb{R}) = \int_{-\infty}^{+\infty} f(x) dx = 1$$

$$F(x) = \int_{-\infty}^x f(u) du$$

II.3.e Caractéristiques

α) Moments

β) Propriétés de $E(X)$ et $\text{var}(X)$

II.3.e.α Moments

Qualifier une distribution par les quantités :

$$\mu_n = \sum p_i x_i^n \quad (\textit{discret})$$

$$\mu_n = \int_{-\infty}^{+\infty} x^n f(x) dx \quad (\textit{continu})$$

appelées **moments**

μ_1 est la **moyenne** ou **espérance mathématique** $E(X)$

$$E(X) = \mu_1 = \sum p_i x_i \quad (\textit{discret})$$

$$E(X) = \mu_1 = \int_{-\infty}^{+\infty} x f(x) dx \quad (\textit{continu})$$

II.3.e.α Moments (suite)

Des moments en déduit les **moments centrés** :

$$\mu_{cn} = \sum p_i (x_i - \mu_1)^n \quad (\textit{discret})$$

$$\mu_{cn} = \int_{-\infty}^{+\infty} (x - \mu_1)^n f(x) dx \quad (\textit{continu})$$

μ_{c2} est la **variance** $\text{var}(X)$, sa racine carrée l'**écart-type** σ

$$\mu_{c2} = \sum p_i (x_i - \mu_1)^2 \quad (\textit{discret})$$

$$\mu_{c2} = \int_{-\infty}^{+\infty} (x - \mu_1)^2 f(x) dx \quad (\textit{continu})$$

II.3.e.α Moments (suite)

On utilise aussi :

μ_{c3} et μ_{c4} dans les coefficients d'asymétrie et d'aplatissement de Fisher

$$\gamma_1 = \frac{\mu_{c3}}{\sigma^3}$$

$$\gamma_2 = \frac{\mu_{c4}}{\sigma^4}$$

II.3.e.β Propriétés de $E(X)$ et $\text{var}(X)$

a et b sont des constantes.

$$E(a) = a$$

$$E(aX + bY) = aE(X) + bE(Y)$$

$$E(XY) = E(X)E(Y) \quad \text{si } X \text{ et } Y \text{ indépendantes}$$

$$\text{var}(a) = 0$$

$$\text{var}(aX + bY) = a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab \text{cov}(X, Y)$$

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

II.3.e.β Propriétés de $E(X)$ et $\text{var}(X)$ (suite)

Cov est un indicateur de liaison dépendant des unités

Pour un indicateur indépendant des unités :

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Corrélation des V.A. X et Y

II.3.f Inégalité de Bienaymé-Tchebichev

μ est l'espérance de la V.A. X

σ son écart-type

λ une constante arbitraire

Alors :

$$P(|X - \mu| \geq \lambda \sigma) \leq \frac{1}{\lambda^2}$$

Cette inégalité limite la probabilité d'un écart à l'espérance (importante pour l'estimation)

Exemples du II.3

- 1) Variable de Bernoulli
- 2) Variable du jeu de dé
- 3) Distribution uniforme
- 4) Variable de survie exponentielle
- 5) Variables centrée et centrée-réduite
- 6) Variable nombre de succès

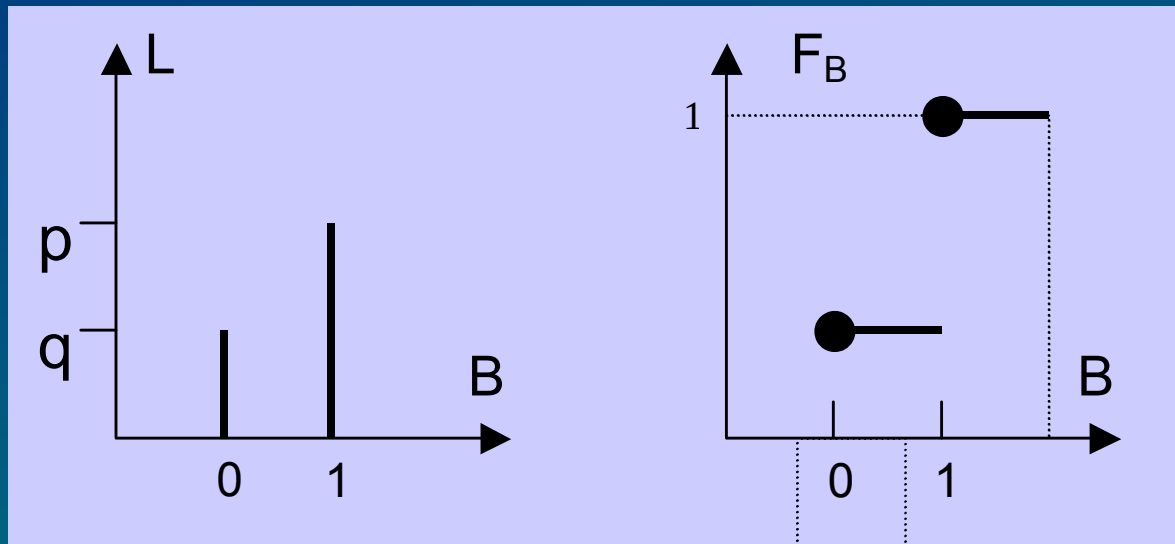
Exemples du II.3 1) Variable de Bernoulli

Expérience à 2 résultats R_a et R_b .

$$\begin{array}{ll} \text{V.A. } B : & R_a \rightarrow 0 \\ & R_b \rightarrow 1 \end{array} \quad \begin{array}{ll} P(R_a) = L(0) = q = 1 - p \\ P(R_b) = L(1) = p = 1 - q \end{array}$$

appelée **variable de Bernoulli**

Loi de proba et fonction de répartition :



Exemples du II.3 1) Variable de Bernoulli (suite)

$$E(B) = 1p + 0q = p$$

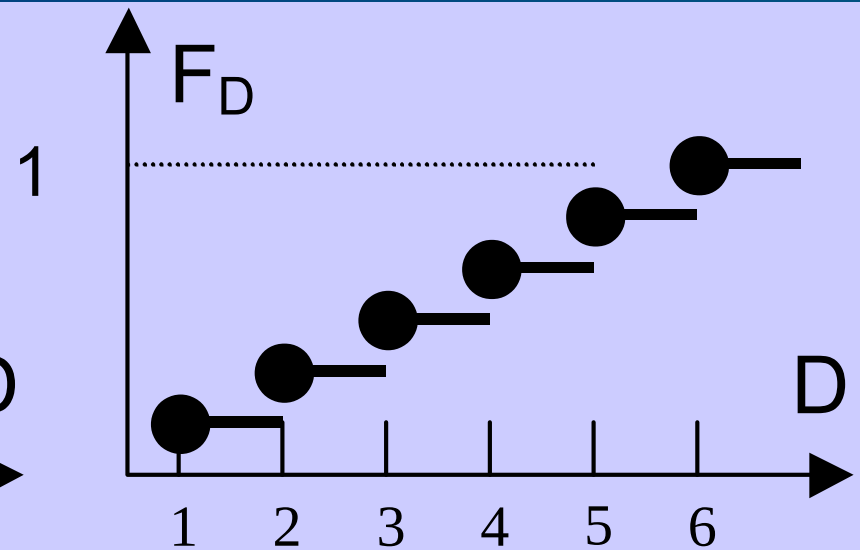
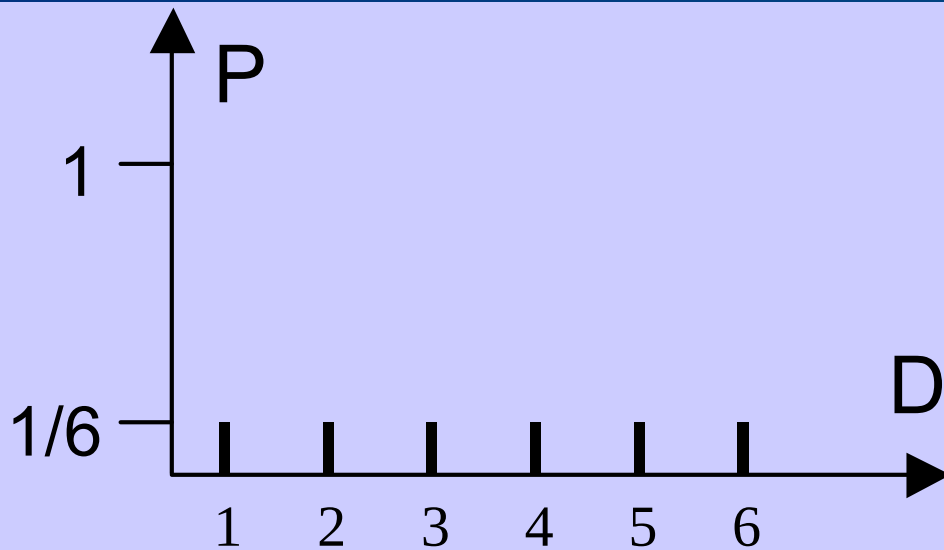
$$\text{Var}(B) = p(1-p)^2 + (1-p)(0-p)^2 = qp$$

Exemples du II.3 2) Variable du jeu de dé

Expérience à 6 résultats F_k .

V.A. $D : F_k \rightarrow k \quad P(F_k) = L(k) = 1/6$

Loi de proba et fonction de répartition :



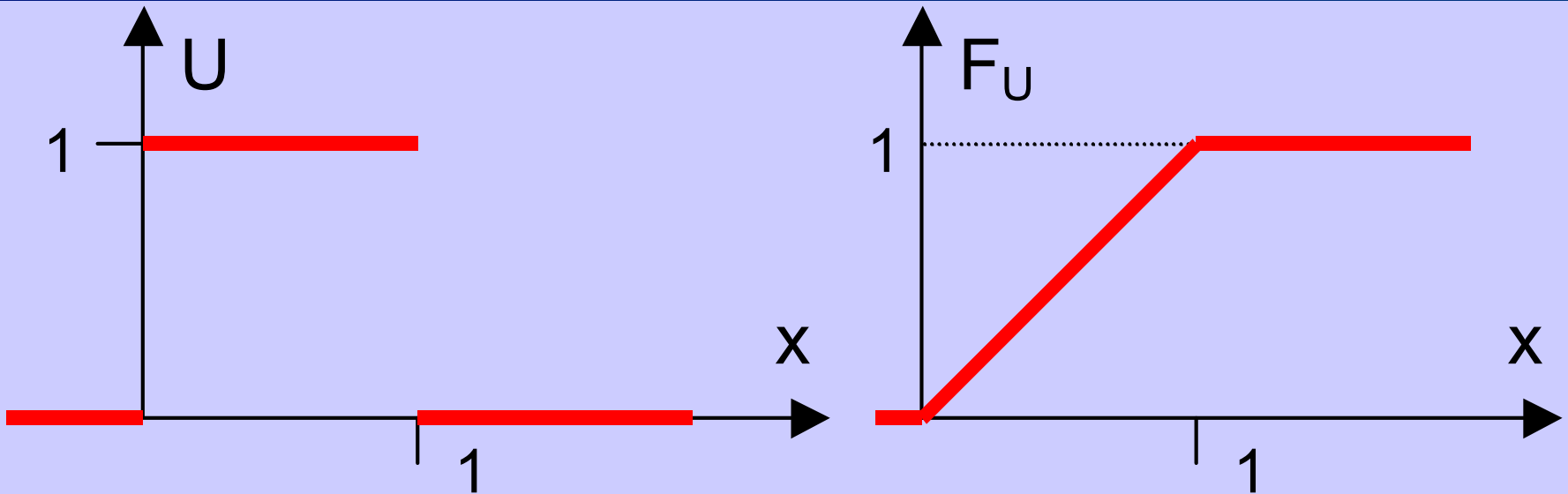
Exemples du II.3 2) Variable du jeu de dé (suite)

$$E(D) = \frac{1}{6} \sum_{i=1}^6 i = 3,5$$

$$\text{var}(D) = \frac{1}{6} \sum_{i=1}^6 (i - 3,5)^2 = 2,917$$

Exemples du II.3 3) Distribution uniforme

Densité $U(x)$ constante sur $[0 ; 1]$



Sert à produire des distributions de nombres aléatoires suivant une V.A. X connaissant la fonction de répartition inverse F_X^{-1} :

$$X = F_X^{-1}(U)$$

Exemples du II.3 4) Variable de survie exponentielle

Expérience aléatoire :

observation de la « survie » d'un atome radioactif.
ensemble E infini non dénombrable des survies

V.A. continue S :

S : survie $\rightarrow s$

s = durée de survie en unités u depuis une origine t_0

Probabilité de désintégration par unité de temps constante :

$$P(\text{désint dans } dt) = k dt$$

Ex. du II.3 4) Variable de survie exponentielle (suite)

$$\begin{aligned}P(S > s + ds) &= P(S > s) P(\text{PAS désint dans } ds) \\&= P(S > s) (1 - k ds)\end{aligned}$$

$$\begin{aligned}P(S > s + ds) - P(S > s) &= -k P(S > s) ds \\dP &= -k P ds\end{aligned}$$

$$dP / P = -k ds$$

Ex. du II.3 4) Variable de survie exponentielle (suite)

Solution :

$$P(S) = C e^{-kS}$$

C une constante à déterminer

La primitive de la solution est la fonction de répartition :

$$F(S) = 1 - (C / k) e^{-kS}$$

Et $F(0) = 0$ donc $C = k$

$$P(S) = k e^{-kS}$$

$$F(S) = 1 - e^{-kS}$$

Ex. du II.3 5) Variables centrée et centrée-réduite

V.A. X donnée.

Deux autres déductibles :

$$X_c = X - E(X)$$

Variable centrée

$$X_{cr} = X_c / \sigma_X = (X - E(X)) / \sigma_X$$

Variable centrée réduite

Ex. II.3 5) Variables centrée et centrée-réduite (suite)

$$E(X_c) = 0$$

$$E(X_{cr}) = 0$$

$$\text{var}(X_{cr}) = 1$$

et l'écart-type aussi !

Ex. du II.3 6) Variable nombre de succès

Expérience :

N répétitions d'une épreuve à issue binaire,
succès **s** ou échec **e**

Résultat :

un N-uplet d'issues **s** ou **e**.

$$E = \{ (r_1, r_2, \dots, r_N) \text{ avec } r_i \in \{ \mathbf{s} ; \mathbf{e} \} \}$$

V.A. « nombre de succès » :

$$K : (r_1, r_2, \dots, r_N) \rightarrow \mathbf{k}$$

k = le nombre de r_i qui sont des **s**.

Ex. du II.3 6) Variable nombre de succès (suite)

des exemples :

$N = 25$ jets de pièce, nombre de piles (succès)

$N = 100$ « tirages d'étudiant », nombre de réussites

$N = 200$ observations de patients, nombre de malades

II.4. Lois de probabilité (usuelles)

a) Lois de variables discrètes

α) Binômiale

β) Hypergéométrique

γ) Poisson

II.4. Lois de probabilité (usuelles)

b) Lois de variables continues

α) Gauss

β) Log-normale (Galton)

γ) Student

δ) Khi-carré (Pearson)

ε) Snedecor

II.4.a.α Loi binômiale

Reprendre l' exemple 6 du II.3

Probabilités des issues d'épreuve :

$$P(\text{s}) = p$$

$$P(\text{e}) = q = 1 - p$$

La loi de K est la **Loi binômiale** (ou de Bernouilli).

elle donne la proba de k succès sur N épreuves **s/e** :

$$B_{N,p}(k) = C_N^k p^k q^{N-k}$$

II.4.a.α Loi binômiale (suite)

Propriétés :

$$E(K) = Np$$

$$\text{var}(K) = Npq$$

$$\gamma_1 = \frac{q-p}{\sqrt{Npq}} \quad \gamma_2 = \frac{1-6pq}{Npq}$$

$$B_{N,p}(k+1) = \frac{N-k}{k+1} \frac{p}{q} B_{N,p}(k)$$

II.4.a.α Loi binômiale (suite)

Au lieu de K , considérer la V.A.

$$F = K / N$$

qui associe la fréquence de succès k / N

F a même loi que K (la binômiale)

Mais :

$$\begin{aligned} E(F) &= p \\ \text{var}(F) &= \frac{pq}{N} \end{aligned}$$

II.4.a.α Loi binômiale (suite)

En appliquant à F l'inégalité de Tchebichev :

Poser e = écart constant :

$$e = \lambda \sigma_F$$
$$P \left(|F - E(F)| \geq e \right) \leq \left(\frac{\sigma_F}{e} \right)^2 = \frac{pq}{e^2 N}$$

F se concentre vers p quand N augmente

(car σ_F diminue quand N croît) **Loi des grands nombres**

Corollaire : augmenter une taille d'échantillon stabilise une fréquence observée.

II.4.a.β Loi hypergéométrique

Urne à N boules, n_1 rouges, n_2 bleues, initialement

Épreuve : tirer une boule sans remise

une boule rouge succès

une boule bleue échec

début : $P(R) = n_1 / N = p$ $P(B) = n_2 / N = q = 1 - p$

Expérience : n répétitions de l'épreuve

Résultat : un n -uplet succès échec avec k succès

$$0 \leq k \leq n \leq N$$

II.4.a.β Loi hypergéométrique (suite)

Loi du nombre de succès (hypergéométrique) :

$$P(K = k) = \frac{C_{n_1}^k C_{n_2}^{n-k}}{C_N^n} = H_{n;p}(k)$$

Analogue de la loi binômiale

mais avec tirage exhaustif (proportions changeantes)

II.4.a.β Loi hypergéométrique (suite)

$$\mu = np$$

$$var = npq \frac{N - n}{N - 1}$$

Quand N devient très grand,
on approche d'une binômiale

II.4.a.γ Loi de Poisson

Mêmes conditions que la loi binômiale, MAIS :

$$N \rightarrow \infty$$

$$p \rightarrow 0$$

Avec :

$$N p = \mu$$

Grand nombre d'essais

Faible chance de succès

Loi de K = Loi de Poisson :

$$P_{\mu}(k) = \frac{e^{-\mu} \mu^k}{k!}$$

II.4.a.γ Loi de Poisson (suite)

Propriétés :

$$E(K) = \mu$$

$$\text{var}(K) = \mu$$

$$P_{\mu}(k+1) = \frac{\mu}{k+1} P_{\mu}(k)$$

Tend vers une loi de Gauss de moyenne et var μ
quand μ croît :

II.4.b.α Loi de Gauss

C'est une densité en forme de cloche dont l'expression est :

$$g(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(\frac{x - \mu}{\sigma \sqrt{2}}\right)^2}$$

Très utile :

- * **Limite** d'autres lois de probabilité (Binômiale, Poisson)
- * Phénomènes aléatoire à **grand nombre de causes additives** indépendantes.

II.4.b.α Loi de Gauss (suite)

Centrée réduite : $\mu = 0$ et $\sigma = 1$

$$g_{\text{CR}}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

C'est celle qui est tabulée

Loi de Gauss = loi normale

II.4.b.α Loi de Gauss (suite)

Propriétés :

$$E(g) = \mu$$

$$\text{var}(g) = \sigma^2$$

$$\gamma_1 = 0 \quad \gamma_2 = 0$$

$$\int_{-\infty}^{+\infty} g(x) dx = 1$$

II.4.b.β Loi log-normale (Galton)

Son log népérien est une loi normale :

$$\text{Ln}(\text{Galton}(x)) = \text{Gauss}(x)$$

si la série (x_i) log normale alors la série $(\text{Ln}(x_i))$ est gaussienne)

Phénomènes à grand nombre de causes multiplicatives
indépendantes.

II.4.b.β Loi log-normale (Galton) (suite)

Paramètres de la loi de Gauss associée :

$$\mu_{\text{Gauss}} = \ln \left(\frac{\mu_{\text{Galton}}^2}{\sqrt{\mu_{\text{Galton}}^2 + \sigma_{\text{Galton}}^2}} \right)$$

$$\sigma_{\text{Gauss}}^2 = \ln \left(1 + \frac{\sigma_{\text{Galton}}^2}{\mu_{\text{Galton}}^2} \right)$$

II.4.b.β Loi log-normale (Galton) (suite)

Formules inverses pour fabriquer une log-normale

$$\mu_{\text{Galton}} = e^{\left(\mu_{\text{Gauss}} + \frac{\sigma_{\text{Gauss}}^2}{2} \right)}$$

$$\sigma_{\text{Galton}}^2 = e^{2 \mu_{\text{Gauss}}} \left(e^{2 \sigma_{\text{Gauss}}^2} - e^{\sigma_{\text{Gauss}}^2} \right)$$

II.4.b.γ Loi de Student

Soit $\nu+1$ v.a. à loi normale X_i (i de 0 à ν)

toutes à $\mu = 0$ et même σ

la v.a. :

$$T = X_0 \sqrt{\frac{\nu}{\sum_{i=1}^{\nu} X_i^2}}$$

Suit la loi de Student

II.4.b.γ Loi de Student (suite)

Forme analogue à Gauss, + aplatie

Dépend du paramètre :

ν = nombre de degrés de liberté (d.d.l.)

$$S(t) = a \left(1 + \frac{t^2}{\nu} \right)^{-\frac{\nu+1}{2}}$$

Converge vers une Gauss quand $\nu > 30$ à 50

Sert à l'estimation et à la comparaison de moyennes avec de petits échantillons

II.4.b.δ Loi de khi-carré (Pearson)

Soit ν v.a. à loi normale X_i (i de 1 à ν)

toutes à $\mu = 0$ et $\sigma = 1$ (centrées réduites)

la v.a. :

$$KHI^2 = \sum_{i=1}^{\nu} X_i^2$$

suit la loi de khi² ou de Pearson

II.4.b.δ Loi de khi-carré (Pearson) (suite)

Densité unimodale, dissymétrique

Dépend du paramètre :

ν = nombre de degrés de liberté

$$P(\chi^2) = a \chi^{\nu-2} e^{-\chi^2/2}$$

Converge lentement vers une Gauss quand ν croît.

Sert à la comparaison de proportions et à l'estimation et la comparaison de variances

II.4.b.ε Loi de Snedecor

Soit KHI^2_1 et KHI^2_2 deux lois de Pearson à ν_1 et ν_2 d.d.l.

La v.a. :

$$F_{\nu_1 ; \nu_2} = \frac{KHI^2_1 / \nu_1}{KHI^2_2 / \nu_2}$$

suit une loi de Snedecor

II.4.b.ε Loi de Snedecor (suite)

Densité unimodale, dissymétrique

Dépend de 2 paramètres (degrés de liberté) :

ν_1 et ν_2

Sert à la comparaison de 2 variances et de multiples moyennes