

# Résumé Probabilités et Statistiques

Lorenzo Brucato

July 2023

## 1 Ensembles, Dénombrement Probabilités évènementielles

### 1.1 Vocabulaire

**Def: Univers.** On désigne par un grand ensemble  $\Omega$  l'univers d'une expérience aléatoire, correspondant à l'ensemble de tous les évènements possibles pour une expérience aléatoire.

**Ex:**

- On lance 6 dés à 6 faces. On utilise un univers  $\Omega$  avec  $6^6$  éléments pour décrire cette expérience (par exemple l'ensemble des 6-uplets  $(v_1, v_2, v_3, v_4, v_5, v_6)$  avec  $v_i \in \{1, 2, 3, 4, 5, 6\}$  qui décrit la valeur de i-ème dé.

- On lance deux dés numérotés de 1 à 6 et on s'intéresse à la somme des deux chiffres. Le résultat est compris dans l'univers  $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

**Def: évènement élémentaire**

Chaque élément élémentaire de  $\Omega$  est appelé **évènement élémentaire**. Un sous ensemble de  $\Omega$  sont souvent appelés **évènements** de l'expérience aléatoire. On a naturellement  $P(\Omega) = 1$

**Notation:** On notera :

- $E_n$  l'ensemble des entiers  $\{1, 2, \dots, n\}$
- $P(A)$ , la probabilité de l'évènement A compris entre 0 (probabilité nulle) et 1 (évènement certain)
- $A^C$ , le complémentaire de l'évènement A (i.e.  $\Omega - A$ )

-  $Card(A) = \#A$ , la taille (ou nombre d'éléments/cardinal) de A.

**Def: Injectif, Surjectif, Bijectif**

- On dit que  $f : A \longrightarrow B$  est **injective** ssi  $\forall (a_1, a_2) \in A$ , si  $a_1 \neq a_2$ , alors  $f(a_1) \neq f(a_2)$ . Lorsque A et B sont de dimensions finies, cela implique par ailleurs  $Card(A) \leq Card(B)$  Autrement-dit, pour chaque élément de l'ensemble de départ, on doit associer un unique élément de l'ensemble d'arrivée.

- On dit que  $f : A \longrightarrow B$  est **surjective** ssi  $\forall y \in B, \exists x \in A : f(x) = y$ . Lorsque A et B sont de dimensions finies, cela implique par ailleurs  $Card(A) \geq Card(B)$  Autrement-dit, pour chaque élément de l'ensemble d'arrivée, il doit exister au moins un antécédant par l'application f.

- On dit que  $f : A \longrightarrow B$  est **bijective** ssi  $f$  est à la fois injective et surjective. Cela revient à dire que pour tout  $y \in B$ , il existe une *unique*  $x \in A$  tel que  $f(x) = y$ . Lorsque A et B sont de dimensions finies, cela implique par ailleurs  $Card(A) = Card(B)$  Autrement-dit, "on forme des couples uniques et distincts" avec un élément de l'ensemble de départ et un élément de l'ensemble d'arrivée.

**Rq:** Pour deux ensembles finis A et B de même cardinal  $n$ , il existe  $n!$  bijections de A dans B possibles.

**Prop:** Pour deux ensembles finis A et B, il existe  $Card(B)^{Card(A)}$  applications possibles de A dans B.

**Def: Ensemble fini, dénombrable, infini**

- On dit que E est un ensemble **fini** si  $Card(E) < \infty$ . Sinon il est **infini**.

- On dit qu'un ensemble infini E est **dénombrable** s'il est en bijection avec  $\mathbb{N}$ . Autrement-dit, on peut "numéroter et compter les éléments de E dans l'ordre".

## 1.2 Coefficient binomial, binome de Newton

**Def: Coefficient binomial**

Le coefficient binomial  $\binom{n}{k}$  est le nombre de combinaisons possibles de k éléments dans un ensemble de n éléments. Il est donné par la formule :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

**Thm: Binome de Newton**

Soient,  $a, b \in \mathbb{R}, n \in \mathbb{N}$  On a :

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

**Prop: Triangle de Pascal**

La formule du triangle de pascal permet de calculer successivement les valeurs du coefficient binomial :

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$$

n\k	0	1	2	3	4	5
1	1	1				
2	1	2	1			
3	1	3	3	1		
4	1	4	6	4	1	
5	1	5	10	10	5	1

**1.3 Probabilités évènementielles, formules de Bayes**

**Def:** Soit  $A, B \in \Omega$  :

- On note  $P(A \cap B)$ , la probabilité de l'intersection des évènements A et B.
- On note  $P(A \cup B)$ , la probabilité de l'union des évènements A et B. On a  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- On note  $P(A|B) = P_B(A)$ , la probabilité de l'évènement A sachant que l'évènement B est réalisé. On a  $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- On dit que A et B sont **indépendants** si l'un des deux évènements n'influe pas sur l'autre se qui se traduit naturellement par l'un des résultats suivants (équivalents par la formule précédente):

$$\begin{aligned} P(A|B) &= P(A), \text{ ou} \\ P(B|A) &= P(B), \text{ ou encore} \\ P(A \cap B) &= P(A).P(B) \end{aligned}$$

**Thm: Formules de Bayes**

Soit  $(\Omega, P)$  une expérience aléatoire d'univers  $\Omega$  et muni d'une mesure de probabilité  $P$ . On a :

$$\begin{aligned} P(A|B) &= P(B|A) \cdot \frac{P(A)}{P(B)} \\ P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)} \end{aligned}$$

**2 Variables aléatoires****2.1 Variables aléatoires réelles (VAR)**

Les variables aléatoires permettent de s'intéresser précisément à des expériences aléatoires ou le résultat peut être décrit par une valeur réelle  $X \in \mathbb{R}$ . On va

s'intéresser ainsi aux probabilités qu'une variable puisse prendre un ensemble de valeurs dans l'ensemble des réels. On notera, pour une mesure de probabilité  $P$ , la probabilité que  $X$  prenne la valeur  $a$  :  $P(X = a)$

**Ex:**

- On s'intéresse au résultat de la somme de deux dés à 6 faces. On peut désigner  $X$  comme étant la variable aléatoire désignant le résultat obtenu.
- On s'intéresse à la taille (en cm) des étudiants d'une promotion. On peut désigner  $X$  comme étant la variable aléatoire valant la taille de l'un des étudiants.

**Def: atome**

Soit  $X$  une VAR. On dit que  $a$  est un **atome** de  $X$  si  $P(X = a) > 0$   
On note  $S_X$  l'ensemble des atomes de  $X$

**Def: VAR Discrète**

Une VAR  $X$  est dite discrète si  $S_X$  est fini ou dénombrable et  $S_X \neq \emptyset, P(X \in S_X) = 1$ .

**Ex:** On désigne par  $X$  la variable aléatoire qui désigne le résultat d'un dé à 6 faces.  $X$  est une VAR discrète puisque  $S_X = \{1, 2, 3, 4, 5, 6\}$  est fini.

**Def: VAR à densité**

Une VAR  $X$  est dite à densité s'il existe une fonction  $f_X$  intégrable ( $F_X(x) = \int_{-\infty}^x f_X(t)dt$ ) et  $S_X = \emptyset$ .

On appelle  $f_X$  la **fonction de densité** de  $X$

La fonction  $F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t)dt$  est la **fonction de répartition** de  $X$  qui désigne la probabilité que  $X$  prenne une valeur inférieure à un certain seuil  $x$ .

**Rq:** Pour une variable à densité, on ne mesure jamais la probabilité en un point (qui est toujours nulle) mais sa densité qui mesure la probabilité dans un voisinage du point.

**Prop:** Soit une VAR  $X$  (discrète ou à densité). Sa fonction de répartition  $F_X$  vérifie les propriétés :

- i)  $F_X$  est croissante, continue à droite
- ii)  $\lim_{t \rightarrow +\infty} F_X(t) = 1$
- iii)  $\lim_{t \rightarrow -\infty} F_X(t) = 0$

## 2.2 Esperance, Variance, Ecart-type

### Def: Esperance

Soit une VAR  $X$ . Si  $X$  est **integrable** (i.e. les formules ci-dessous ont un sens, résultat fini en valeur absolue, cf cours d'integration), on définit **l'esperance** de  $X$  (ou premier moment) comme étant la valeur :

$$E[X] = \sum_{k \in S_X} k.P(X = k), \text{ si } X \text{ est discrète}$$
$$E[X] = \int_{\mathbb{R}} t.f_X(t)dt, \text{ si } X \text{ est à densité}$$

La loi des grands nombres (cf. partie 3) permet de comprendre l'esperance comme "la moyenne" des résultats obtenus pour des réalisations (infinies) successives de  $X$  (ou la valeur moyenne espérée).

### Prop: Formule de transfert

Soit  $X$  une VAR,  $f$  une fonction définie sur  $\mathbb{R}$ . On a :

$$E[f(X)] = \sum_{k \in S_X} f(k).P(X = k), \text{ si } X \text{ est discrète}$$
$$E[f(X)] = \int_{\mathbb{R}} f(t).f_X(t)dt, \text{ si } X \text{ est à densité}$$

### Def: Variance

Soit une VAR  $X$ . On définit **la variance** de  $X$  comme étant la valeur :

$$Var(X) = E[X^2] - E[X]^2 = E[(X - E[X])^2]$$

On note l'écart-type  $\sigma = \sqrt{Var(X)}$

La variance est l'esperance de la variable aléatoire décrivant les écarts à la moyenne de  $X$ . Une variance élevée indique une forte fluctuation des résultats autour de l'esperance. Une variance faible indique une concentration forte des réalisations de  $X$  près de l'esperance.

## 2.3 Moments, Fonction génératrice des moments

### Def: Moment d'ordre k

On appelle moment d'ordre  $k$  l'esperance de  $X^k$  (si intégrable).

### Def: Fonction génératrice des moments

On appelle fonction génératrice des moments la fonction  $G_X(t) = E[e^t X] = \sum_{k=0}^{\infty} E[X^k] \frac{t^k}{k!}$ . Elle permet en particulier de retrouver efficacement les moments de  $X$ .

## 2.4 Lois discrètes

On s'intéresse ici aux lois classiques que peuvent suivre une variable aléatoire discrète.

### 2.4.1 Loi de Dirac

Loi dont toute la masse de probabilité est concentrée au point  $a$  (La réalisation de  $X \sim D(a)$  est de 1.0 en  $a$ )

**Prop:**

- $E[X] = a$
- $Var(X) = 0$

### 2.4.2 Loi Uniforme discrète

$X \sim Unif(S_X)$  modélise une expérience aléatoire sur un nombre fini d'issues équiprobables.

**Prop:**

- $E[X] = \frac{1}{Card(S_x)} \sum_{x \in S_x} x$
- $Var(X) = \frac{Card(S_x)^2 - 1}{12}$

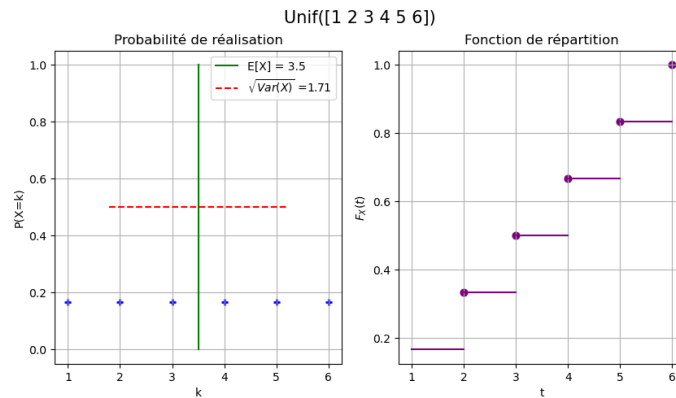


Figure 1: Loi uniforme discrète

### 2.4.3 Loi de Bernoulli

$X \sim \text{Ber}(p)$  modélise une expérience aléatoire à deux issues 0 (échec), 1 (succès), avec probabilité  $p$  de succès.

**Prop:**

- $E[X] = p$
- $\text{Var}(X) = p(1 - p)$

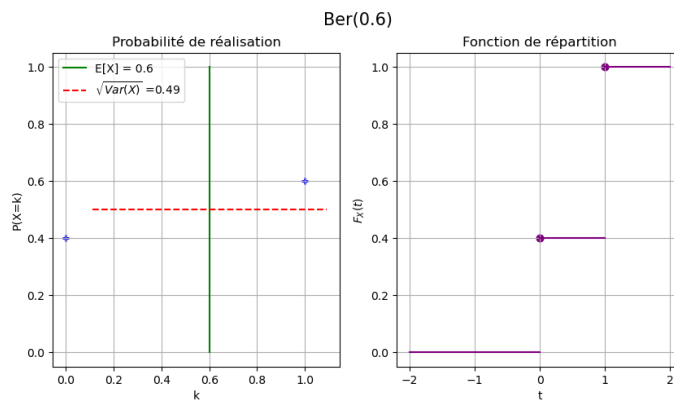


Figure 2: Loi de Bernoulli

### 2.4.4 Loi Binomiale

$X \sim \text{Bin}(n, p)$  modélise la répétition de  $n$  expériences de Bernoulli identiques et indépendantes de paramètre  $p$ .

**Prop:**

- $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
- $E[X] = np$
- $\text{Var}(X) = np(1 - p)$
- $G_X(t) = (pe^t + 1 - p)^n$

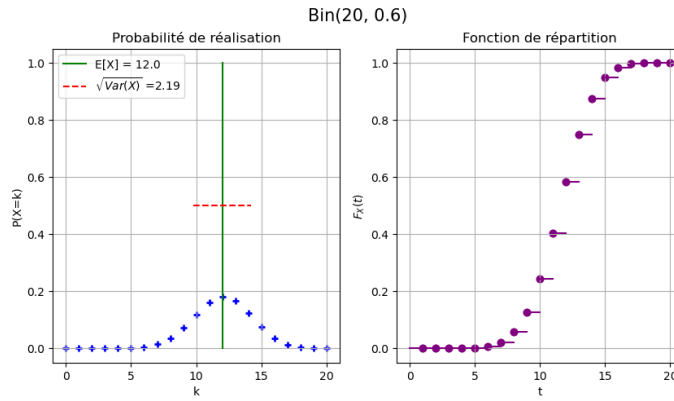


Figure 3: Loi Binomiale

#### 2.4.5 Loi de Poisson (ou des évènements rares)

$X \sim \text{Poi}(a)$  modélise une expérience aléatoire dont la probabilité de réalisation diminue de manière très significative (évènements rares).

**Prop:**

- $P(X = k) = e^{-a} \frac{a^k}{k!}$
- $E[X] = a$
- $\text{Var}(X) = a$
- $G_X(t) = e^{a(e^t - 1)}$

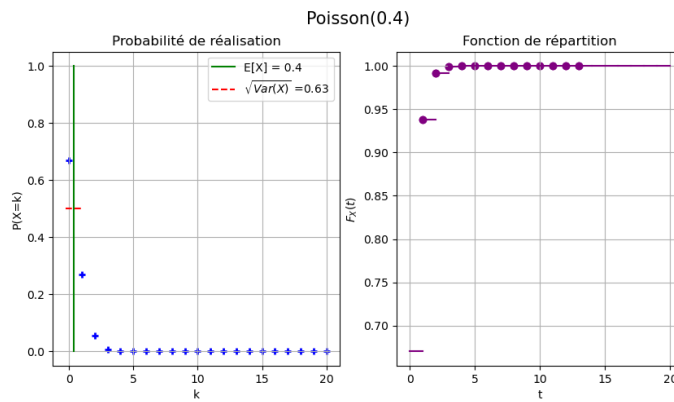


Figure 4: Loi de Poisson



### 2.4.6 Loi Géométrique

$X \sim \text{Geom}(p)$  modélise une expérience aléatoire qui s'intéresse au nombre de réalisations avant succès d'une expérience aléatoire à deux issues

**Prop:**

- $P(X = k) = p(1 - p)^{k-1}$
- $E[X] = 1/p$
- $\text{Var}(X) = \frac{1-p}{p^2}$

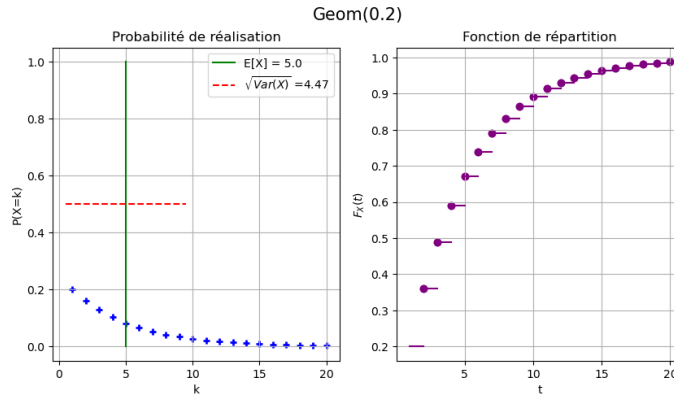


Figure 5: Loi Binomiale

### 2.4.7 Loi Hypergéométrique

$X \sim H(n, N, m)$  modélise une expérience aléatoire qui s'intéresse au nombre de boules blanches tirées sur  $n$  tirages dans une urne à  $N$  boules contenant  $m$  boules blanches

**Prop:**

$$- P(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}$$

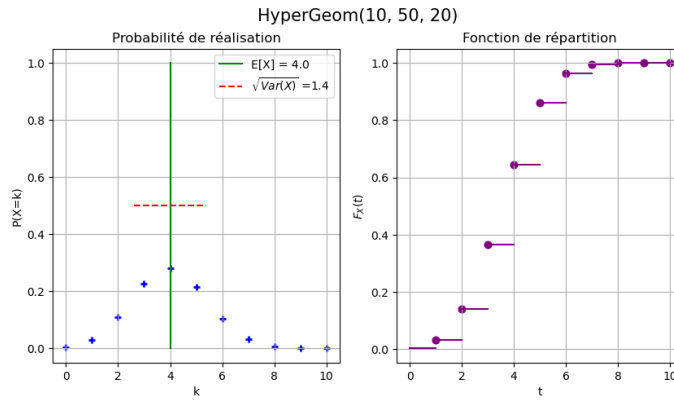


Figure 6: Loi Hypergéométrique

## 2.5 Lois à densité

Quelques lois usuelles des variables aléatoires à densité.

### 2.5.1 Loi Uniforme à densité

$X \sim \text{Unif}(I = [a, b])$ , I intervalle continu, modélise une expérience aléatoire équiprobable sur l'intervalle I.

**Prop:**

- $f_X(x) = \frac{1_I(x)}{\text{Card}(I)}$
- $F_X(t) = \frac{t-a}{b-a}1_I(t) + 1_{t>b}(t)$
- $E[X] = (a+b)/2$
- $\text{Var}(X) = (b-a)^2/12$

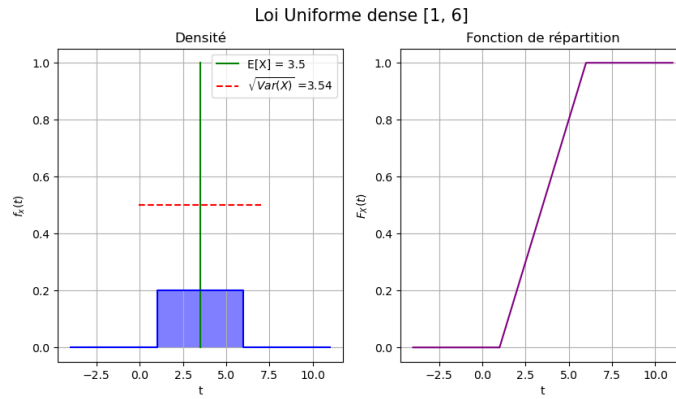


Figure 7: Loi Uniforme dense

### 2.5.2 Loi Exponentielle

$X \sim \text{Exp}(\lambda)$  modélise une expérience aléatoire dont la probabilité de réalisation décroît exponentiellement.

**Prop:**

- $f_X(x) = \lambda e^{-\lambda x} 1_{x \geq 0}$
- $F_X(t) = (1 - e^{-\lambda t}) 1_{t \geq 0}$
- $E[X] = 1/\lambda$
- $\text{Var}(X) = 1/(\lambda^2)$
- $G_X(t) = \lambda/(\lambda - t)$

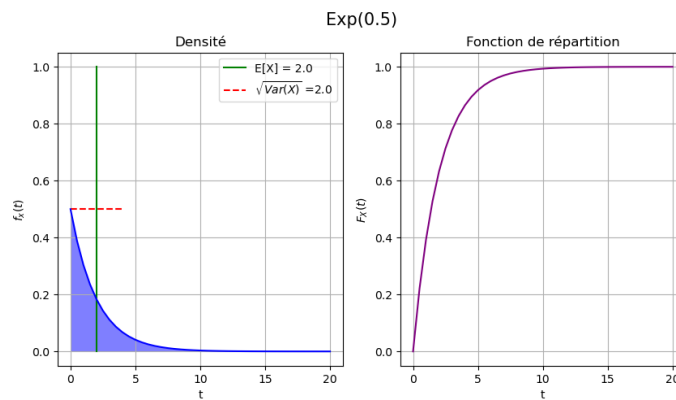


Figure 8: Loi Exponentielle

### 2.5.3 Loi Normale (Gaussienne)

$X \sim \text{Exp}(\mu, \theta^2)$  modélise la distribution classique autour de l'esperance  $\mu$  donnée par le théorème central limite. La loi  $N(0, 1)$  d'esperance nulle et de variance 1 est *la loi normale centrée réduite*.

**Prop:**

- $f_X(x) = \frac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{(x-\mu)^2}{2\theta^2}}$
- $F_X(t) = \int_{-\infty}^t f_X(x) dx$
- $E[X] = \mu$
- $\text{Var}(X) = \theta^2$
- $G_X(t) = e^{t\mu + (\theta^2 t^2)/2}$

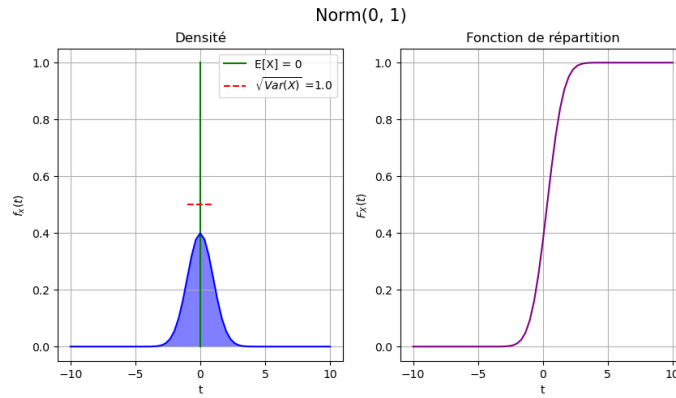


Figure 9: Loi Normale

### 2.5.4 Loi Gamma

La loi Gamma  $X \sim \text{Gamma}(\alpha, \beta)$  est la loi générale suivie par une somme de variables suivants une loi exponentielle.  $\text{Exp}(\lambda)$  est le cas particulier où  $\alpha = 1$ ,  $\beta = \lambda$

**Prop:**

- $\Phi(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-\beta x} dx$
- $f_X(x) = \beta^\alpha \Phi(\alpha)^{-1} x^{\alpha-1} e^{-\beta x} 1_{x \geq 0}$
- $F_X(t) = \int_{-\infty}^t f_X(x) dx$
- $E[x] = \alpha/\beta$
- $\text{Var}(X) = \alpha/\beta^2$

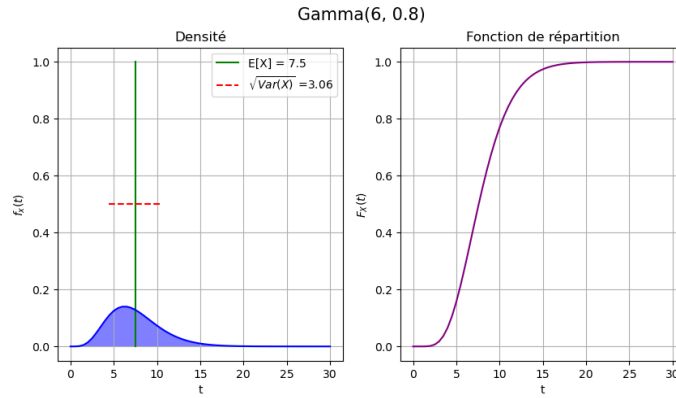


Figure 10: Loi Gamma

## 2.6 Loi centrée réduite et Theoreme de Moivre Laplace

### Def : Loi centrée réduite

Soit  $X$  une VAR de loi paramétrique donnée. On appelle loi centrée réduite de  $X$ , la variable aléatoire :

$$Y = \frac{X - E[X]}{\sqrt{\text{Var}(X)}}$$

Son espérance vaut 0 et sa variance vaut 1.

### Thm: Moivre-Laplace

Une variable aléatoire  $X \sim \text{Bin}(n, p)$  peut être approchée par la loi gaussienne  $N(np, np(1-p))$  lorsque  $np(1-p) \gg 10$ . Il s'agit d'un résultat du théorème central-limite (cf. 3. Statistiques)

La loi normale donne ainsi une bonne approximation de la loi binomiale lorsque  $n$  tend vers l'infini et  $p \sim 0.5$

### Prop: approximation par loi de Poisson

De manière similaire, une variable aléatoire  $X \sim \text{Bin}(n, p)$  peut être approchée par la loi de Poisson  $\text{Poi}(p)$  lorsque  $p$  faible et  $n$  tend vers l'infini.

## 2.7 Inégalités de Markov, Chebychev et Jensen

### Thm: Inégalité de Markov (Inégalité sur l'espérance)

Soit une VAR  $X$  positive intégrable. On a :

$$\forall x \geq 0, P(X \geq x) \leq \frac{E[X]}{x}$$

### Thm: Inégalité de Chebychev (Inégalité sur la variance)

Soit une VAR  $X$  de carré intégrable. On a :

$$\forall x \geq 0, P(|X - E[X]| \geq x) \leq \frac{Var(X)}{x^2}$$

### Thm: Inégalité de Jensen

Soit  $\phi : I \rightarrow \mathbb{R}$  convexe,  $I$  ouvert.

Soit une VAR  $X$  intégrable sur  $I$  et  $P(X \in I) = 1$ . On a :

$$\phi(E[X]) \leq E[\phi(X)]$$

## 2.8 Vecteur aléatoire

### Def: Vecteur aléatoire

On appelle vecteur aléatoire  $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ , un vecteur de  $n$  variables aléatoires  $X_1, \dots, X_n$ . Sa fonction de densité  $f_X$  est une fonction de  $\mathbb{R}^n$  dans  $\mathbb{R}$

### Def: densité marginale

Chaque composante  $X_i$  d'un vecteur aléatoire  $X$  a pour densité marginale :

$$f_{X_i}(x) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_X(x_1, \dots, x, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

Si  $X$  est à densité, alors toutes les composantes  $X_i$  sont à densité.

## 2.9 Propriétés d'indépendance et changement de variable

### Def: Indépendance de VAR

Soit  $X \in \mathbb{R}^n$ .  $X_1, \dots, X_n$  sont indépendantes si et seulement si, de manière équivalente :

- i)  $P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i)$
- ii)  $F_{X_1 \dots X_n} = P(X_1 \leq x_1) \dots P(X_n \leq x_n)$
- iii) Si  $X$  discrete,  $P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$   
Si  $X$  à densité,  $f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$

### Prop: propriétés d'indépendance

Si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires indépendantes, alors :

$$Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i)$$

$$E[\prod_{i=1}^n g_i(X_i)] = \prod_{i=1}^n E[g_i(X_i)]$$

$$G_{X_1+\dots+X_n} = \prod_{i=1}^n G_{X_i}(t)$$

**Prop: densité de somme de variables aléatoires**

On appelle convolée de  $f$  et  $g$  :

$$(f * g)(x) = \int_{\mathbb{R}} f(t)g(x-t)dt$$

Si  $X_1, \dots, X_n$  sont des VAR indépendantes de densités respectives  $f_{X_1}, \dots, f_{X_n}$ , alors  $X_1 + \dots + X_n$  est de densité  $f_{X_1} * \dots * f_{X_n}$

**Thm: Changement de variable**

Soit  $\Phi : A \rightarrow B$ , une bijection différentiable continue et sa réciproque  $\Phi^{-1} : B \rightarrow A$ .

On note  $J$  la jacobienne de  $\Phi$ .

Alors,  $\forall g : \mathbb{R}^n \rightarrow \mathbb{R}$  bornée et  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  intégrable, on a:

$$\int_B g(\Phi(x))f(x)dx = \int_A g(y)f(\Phi^{-1}(y))|J(\Phi^{-1}(y))|dy$$

## 2.10 Convergence de lois

On s'intéresse à la convergence, quand  $n$  tend vers l'infini, d'un vecteur  $X_n$  de  $n$  variables aléatoires indépendantes et identiquement distribuées.

**Def : Convergence en loi**

On dit que  $X_n$  converge en loi vers  $X$ ,  $X_n \xrightarrow{loi} X$  si :

$$\lim_{n \rightarrow +\infty} P(X_n \leq t) = P(X \leq t)$$

**Def : Convergence en probabilité**

On dit que  $X_n$  converge en probabilité vers  $X$ ,  $X_n \xrightarrow{P} X$  si :

$$\forall \epsilon > 0, \lim_{n \rightarrow +\infty} P(|X_n - X| > \epsilon) = 0$$

La convergence en probabilité implique la convergence en loi

**Def : Convergence  $L^p$**

On dit que  $X_n$  converge en  $L^p$  vers  $X$ ,  $X_n \xrightarrow{L^p} X$  si :

$$\forall p > 0, \lim_{n \rightarrow +\infty} E[|X_n - X|^p] = 0$$

La convergence en  $L^p$  implique la convergence en probabilité

## 3 Statistiques

### 3.1 Echantillon de données et estimateur

**Def: loi paramétrique**

Une loi paramétrique est une loi d'un ou plusieurs paramètres réels  $\theta_1, \theta_2, \dots$  (Par exemple l'esperance et la variance pour la loi gaussienne...)

**Def: n-échantillon**

On appelle **n-échantillon**, un vecteur de  $n$  variables aléatoires supposées de même loi et indépendantes (**iid**). Elles sont issues d'une expérience aléatoire sur lequel on veut estimer certains paramètres (par exemple : n-échantillon de médicaments pour un test de fiabilité, n-échantillon de pièces mécaniques pour test de durabilité, n-échantillon de personnes pour un sondage, ...)

**Def: Estimateur**

Soit un n-échantillon  $X_1, \dots, X_n$  de  $n$  variables aléatoires iid de loi paramétrique donnée et  $\theta$  un paramètre de la loi. Un **estimateur**  $\hat{\theta}$  de  $\theta$  est une variable aléatoire fonction du n-échantillon  $X_1, \dots, X_n$ , construit dans le but d'estimer  $\theta$ .

**Def: Biais d'un estimateur**

Le biais d'un estimateur  $\hat{\theta}$  pour estimer  $\theta$  est donné par :

$$B(\hat{\theta}, \theta) = E[\hat{\theta}] - \theta$$

Il s'agit de l'écart moyen entre l'esperance de notre estimateur et la valeur réelle du paramètre.

On dit que  $\hat{\theta}$  est **sans biais** si  $B(\hat{\theta}, \theta) = 0$

**Def: Risque quadratique**

Le risque quadratique est basé sur l'esperance de l'écart entre un estimateur  $\hat{\theta}$  de  $\theta$  et la valeur réelle du paramètre  $\theta$ . Il permet ainsi d'évaluer la qualité d'un estimateur. Il est donné par :

$$R(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + B(\hat{\theta}, \theta)^2$$

**Def: Consistance d'un estimateur**

Un estimateur  $\hat{\theta}$  est dit **consistant** si, lorsque la taille de l'échantillon augmente, il converge vers la vraie valeur du paramètre  $\theta$ , ce qui se traduit, de manière équivalente, par :

i)  $\lim_{n \rightarrow +\infty} R(\hat{\theta}) = 0$

ii)  $\hat{\theta} \xrightarrow{P} \theta$



### 3.2 Estimateur des moments

**Thm: Loi des grands nombres**

Soit  $X_1, \dots, X_n$  un n-échantillon. La moyenne empirique, notée  $\bar{X} = (X_1 + \dots + X_n)/n$ , converge en probabilité vers  $E[X_1]$

**Def: Estimateur des moments**

La linéarité de l'espérance et le résultat de la loi des grands nombres permettent d'utiliser la moyenne empirique comme estimateur sans biais et consistant des moments des  $X_i$ .

### 3.3 Estimateur du maximum de vraisemblance

**Def: Vraisemblance**

Soit un n-échantillon  $X_1, \dots, X_n$  de densité  $f_\theta$ , dépendant du paramètre  $\theta$ . La **vraisemblance** de l'échantillon est donnée par :

$$L_n(\theta) = \prod_{i=1}^n f_\theta(X_i)$$

**Def: Estimateur du maximum de vraisemblance**

Rechercher le maximum de  $L_n(\theta)$  revient à rechercher l'estimateur de  $\theta$  qui rend le plus vraisemblable la réalisation du n-échantillon. S'il existe, on construit ainsi un nouvel estimateur :

$$\hat{\theta}_{MV} = \max(L_n(\theta))$$

On peut, puisque que le logarithme est croissant, choisir un estimateur parfois plus simple à déterminer :

$$\hat{\theta}_{MV} = \max(\log L_n(\theta))$$

### 3.4 Inégalité de Cramer-Rao

La question est désormais de savoir : jusqu'à quel point un estimateur peut-il être "performant" ?

**Def: Information de Fisher**

On définit l'information de Fisher par :

$$I_n(\theta) = \text{Var}(\log L'_n(\theta))$$

**Thm : Borne de Cramer-Rao**

Dans un modèle dit "régulier" (notion hors programme, mais qui concernera l'ensemble des cas étudiés) :

Soit  $g$  une fonction dérivable sur le domaine de  $\theta$ ,  $\hat{\theta}$  estimateur sans biais de  $g(\theta)$ , on a :

$$R(\hat{\theta}) \geq \frac{g'(\theta)^2}{I_n(\theta)}$$

Cette inégalité montre en particulier, que tout estimateur possède un risque minimum...

**3.5 Intervalles de confiance asymptotique****Thm: Theoreme central limite**

Soit  $X_n$  une suite de VAR iid de carré intégrable. On a alors :

$$\sqrt{n}(\bar{X} - E[X_1]) \xrightarrow{loi} N(0, Var(X_1))$$

**Lemme: Lemme de l'application continue**

Soit  $g$  une fonction continue. Si  $Z_n$  converge en loi (resp. probabilité) vers  $Z$ , alors  $g(Z_n)$  converge en loi (resp. probabilité) vers  $g(Z)$

**Thm: Slutsky**

Si  $(X_n)$  converge en loi vers  $X$ , et  $(Y_n)$  converge en loi vers une constante réelle  $c$  (Dirac( $c$ )), alors le couple  $(X_n, Y_n)$  converge en loi vers  $(X, c)$

**Prop: Methode delta**

Soit  $X_n$  une suite de VAR iid de carré intégrable et  $g$  une fonction dérivable,  $g' \neq 0$  On a alors :

$$\sqrt{n}(g(\bar{X}) - g(E[X_1])) \xrightarrow{loi} N(0, g'(E[X_1])^2 Var(X_1))$$

**Def : Estimateur asymptotiquement gaussien**

On dit qu'un estimateur  $\hat{\theta}$  de  $\theta$  est asymptotiquement gaussien si:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{loi} N(0, \sigma^2)$$

On peut construire à l'aide de ce résultat et en utilisant les quantiles de la loi normale, un intervalle de confiance asymptotique pour estimer  $\theta$ , avec une erreur  $\alpha$  que l'on choisira. De plus, puisque le résultat est asymptotique, il ne devient pertinent qu'au fur et à mesure que la taille de l'échantillon augmente...

### 3.6 Intervalles de confiance de la loi normale

On considère dans cette partie un n-échantillon  $X_n$  suivant une loi gaussienne  $N(\mu, \sigma^2)$ . On peut dans ce cas construire des intervalles de confiance non asymptotiques pour estimer soit  $\mu$ , soit  $\sigma^2$  :

**Prop : estimer  $\mu$  si  $\sigma^2$  connu**

Si  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ , alors  $\bar{X} \sim N(\mu, \sigma^2/n)$  est un estimateur de  $\mu$  et permet de construire un intervalle de confiance pour  $\mu$ .

**Prop : estimer  $\mu$  si  $\sigma^2$  inconnu - loi de Student**

Soit  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ .

On pose  $s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$ , l'estimateur de la variance empirique.

On note  $T(n)$  la **loi de Student**, non explicitée ici, mais dont on connaît une table des valeurs.

Alors :

$$\sqrt{n} \frac{\bar{X} - \mu}{s} \sim T(n-1)$$

Cela permet de construire un intervalle de confiance pour  $\mu$

**Prop : estimer  $\sigma^2$  - loi du Khi-2**

On note  $X_2(n)$  la **loi du Khi-2**, non explicitée ici, mais dont on connaît une table des valeurs. On a les deux propriétés :

$$- (1/\sigma) \sum (X_i - \mu)^2 \sim X_2(n)$$

Ce résultat permet d'estimer  $\sigma$  si  $\mu$  est connu et de construire un intervalle de confiance pour  $\sigma$

$$- (1/\sigma) \sum (X_i - \bar{X})^2 \sim X_2(n-1)$$

Ce résultat permet d'estimer  $\sigma$  si  $\mu$  est inconnu et de construire un intervalle de confiance pour  $\sigma$

### 3.7 Methodologie de tests

On considèrera dans cette partie un n-échantillon  $X_n$  suivant une loi paramétrique  $P(\theta)$ . On réalise un test en opposant deux hypothèses l'une contre l'autre :

- l'hypothèse nulle  $H_0 : \theta \in \Theta_0$
- l'hypothèse alternative  $H_1 : \theta \in \Theta_1$

On définit ensuite une **région de rejet**  $R$ . Il s'agit de l'ensemble des réalisations de  $X_n$  pour lesquelles on décide de rejeter  $H_0$  au profit de  $H_1$ .

**Def: statistique de test**

On définit plusieurs stratégies de tests :

- un **test unilatéral** consiste à rejeter  $H_0$  lorsque la valeur  $T(X_n)$  dépasse un certain seuil  $c$  (soit positivement, soit négativement).
- un **test bilatéral** consiste à rejeter  $H_0$  lorsque la valeur  $|T(X_n)|$  dépasse un certain seuil  $c$ . (à la fois positivement et négativement).

**Def:**

i) On appelle **erreur de première espèce** (erreur la plus grave), la probabilité de rejeter  $H_0$  alors que  $H_0$  est vraie. Plus précisément c'est la probabilité, sachant  $H_0$ , que l'évènement  $\theta \in R$  se réalise :

$$\alpha(\theta) = P_{H_0}((X_1, \dots, X_n) \in R)$$

ii) On appelle **erreur de seconde espèce** (erreur moins grave), la probabilité de conserver  $H_0$  alors que  $H_0$  est fautive. Plus précisément, c'est la probabilité, sachant  $H_1$ , que l'évènement  $\theta \notin R$  se réalise :

$$\beta(\theta) = P_{H_1}((X_1, \dots, X_n) \notin R)$$

**Def: taille du test**

On définit la taille du test :  $t = \sup_{\theta \in \Theta_0} \alpha(\theta)$

On dit que le test est de niveau de confiance  $a$  si  $a \geq \sup_{\theta \in \Theta_0} \alpha(\theta)$ .

**Def: puissance du test**

La puissance du test est sa capacité à rejeter  $H_0$  lorsque  $H_0$  est effectivement faux :

$$\pi(\theta) = 1 - \beta(\theta) = P_{H_1}(X_n \in R)$$

**Rq:**

Si  $I_n$  est un intervalle de confiance de niveau  $1 - a$  pour estimer  $\theta$ , alors le test de rejet  $R = \{(X_1, \dots, X_n), \theta_0 \notin I_n\}$  est un test de niveau de confiance  $a$  pour tester  $H_0 : \theta_0 = \theta$  contre  $H_1 : \theta_0 \neq \theta$

**Def:**

On dit qu'un test  $\Phi_1$  est **uniformément plus puissant** (UPP) que  $\Phi_2$  si  $\pi_{\Phi_1}(\theta) \geq \pi_{\Phi_2}(\theta) \forall \theta \in \Theta_1$

**Lemme: Neyman-Pearson**

Si l'on teste une hypothèse  $H_0 : \theta = \theta_0$  contre  $H_1 : \theta = \theta_1$  ou encore  $H_0 : f_\theta = f_{\theta_0}$  contre  $H_1 : f_\theta = f_{\theta_1}$ , alors le test de rejet :

$$R = \{\prod f_1(X_i) / \prod f_0(X_i) > c_a\}$$

de taille  $a$  est UPP que tout autre test de niveau de confiance au plus  $a$ .

**Def: p-value**

On appelle p-valeur (p-value en anglais), la probabilité dans l'hypothèse  $H_0$  que l'on ait une valeur "plus extrême" que celle observée :

- pour un test unilatéral à droite :  $p_{val} = P(X_n > X^{obs})$
- pour un test unilatéral à gauche :  $p_{val} = P(X_n < X^{obs})$
- pour un test bilatéral :  $p_{val} = P(|X_n| > |X^{obs}|)$