Part. 1.1. Bases et intuition de la théorie des probabilités

K. Meziani, IASD

Dauphine | PSL
EXECUTIVE EDUCATION

Table des matières

Plan

Supports complémentaires

https://sites.google.com/site/katiameziani00/home/enseignements

- Polycopié de probabilités (Partie 1)
- Polycopié de statistiques descriptives (Partie 2)

1. Variable aléatoire

Variable aléatoire discrète X

- Le support $Supp_X$ de X est l'ensemble des valeurs possibles x prises par X.
- Le cardinal du support de X est fini ou au plus dénombrable.

Exemple: résultat du lancé d'un dé, nombre d'accidents dans l'année, nombre de personnes dans une file d'attente...

Variable aléatoire continue X

- Le support $Supp_X$ de X est inclus dans $\mathbb{R}$.
- Le cardinal du support de X est infini.

Exemple: taille d'une personne, position d'une particule....

Probabilités

Probabilité:

mesure de valeurs possibles x, prises par la variable aléatoire X.

Exemples :

- $P(X = 2)$, $P(X \leq 2)$, $P(1 < X \leq 5)$ ...
- Pour les variables aléatoires continues les probabilités ponctuelles nulles : $\forall x \in Supp_X$, $P(X = x) = 0$.

Comment décrire la Loi de probabilité de X

Distribution de probabilité

$$\{P(X = x)\}_{x \in Supp_X}$$

- Pour tout $x \in Supp_X$

$$0 < P(X = x) \leq 1$$

- Somme des probabilités sur le support vaut 1:

$$\sum_{x \in Supp_X} P(X = x) = 1.$$

- **Uniquement pour les v.a. discrètes**

Densité de probabilité

$$f(x)$$

- Pour tout $x \in \mathbb{R}$,

$$f(x) \geq 0$$

- Équivalent de la somme des probabiliés cas continue

$$\int_{-\infty}^{+\infty} f(t)dt = 1.$$

- f est continue presque partout,

$$Supp_X = \{x \in \mathbb{R}, f(x) > 0\}$$

Espérance, variance, écart type

V.a. discrète

$$\{P(X=x)\}_{x \in Supp_X}$$

- Espérance/moyenne

$$E[X] = \sum_{supp_X} xP(X=x)$$

- Variance, si $E[X] < \infty$

$$V[X] = \sum_{supp_X} (x - E[X])^2 P(X=x).$$

- Écart type

$$\text{Écart type} = \sqrt{V[X]}$$

V.a. continue

$$f(x)$$

- Espérance/moyenne

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx$$

- Variance, si $E[X] < \infty$

$$V[X] = \int_{-\infty}^{+\infty} (x - E[X])^2 f(x)dx.$$

- Écart type

$$\text{Écart type} = \sqrt{V[X]}$$

Lois usuelles discrètes

Loi	Support	$P(X=k)$	Espérance	Variance
Uniforme $\mathcal{U}(\{1,\cdots,n\})$	$\{1,\cdots,n\}$	$1/n$	$\frac{(n+1)}{2}$	$\frac{(n+1)(n-1)}{12}$
Bernoulli $\mathcal{B}(p)$ $p\in]0,1[$	$\{0,1\}$	$P(X=1)=p$ $P(X=0)=1-p$	p	$p(1-p)$
Binomiale $\mathcal{B}in(n,p)$ $n\in\mathbb{N}^*$ $p\in]0,1[$	$\{0,\cdots,n\}$	$C_n^k p^k(1-p)^{n-k}$	np	$np(1-p)$
Poisson $\mathcal{P}(\lambda)$ $\lambda>0$	$\mathbb{N}$	$\lambda^k e^{-\lambda}/k!$	λ	λ
Géométrique $\mathcal{G}(p)$ $p\in]0,1[$	$\mathbb{N}^*$	$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Lois usuelles continues

Loi	Supp.	Densité	Espér.	Var.
Uniforme $\mathcal{U}([a,b])$ $(a,b) \in \mathbb{R}^2$, $a < b$	$[a,b]$	$f(x) = \frac{1}{b-a}$ si $x \in [a,b]$ $= 0$ sinon	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponentielle $\mathcal{E}(\lambda)$ $\lambda > 0$	$\mathbb{R}_+^*$	$f(x) = \lambda e^{-\lambda x}$ si $x > 0$ $= 0$ sinon	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normale $\mathcal{N}(0,1)$	$\mathbb{R}$	$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$	0	1
Normale $\mathcal{N}(\mu,\sigma)$ $(\mu,\sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\mathbb{R}$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	μ	σ^2
Gamma $\mathcal{G}(k,\theta)$ $(k,\theta) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$	$\mathbb{R}_+^*$	$f(x) = \frac{\theta^k}{\Gamma(k)} x^{k-1} e^{-\theta x}$ si $x > 0$ $= 0$ sinon	k/θ	k/θ^2

Loi uniforme continue sur $[a, b]$

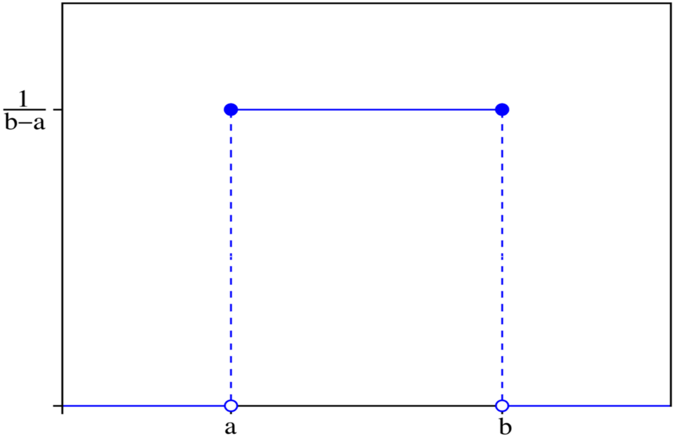

Figure 1: Densité de loi uniforme

Lois Exponentielles

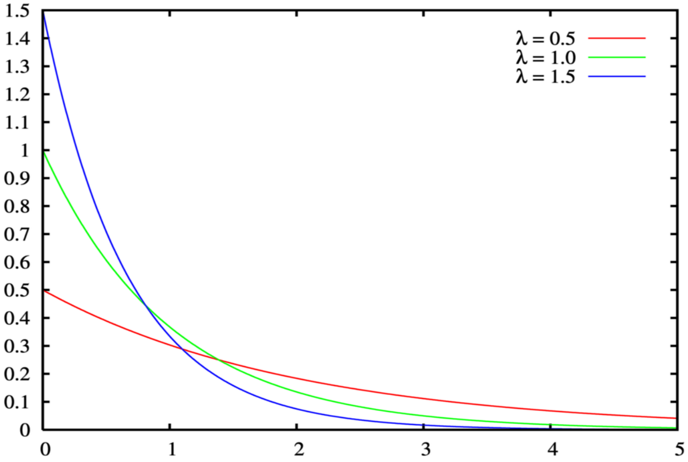

Figure 2: Densités de lois exponentielles pour différents paramètres

Lois Gaussiennes/Normales

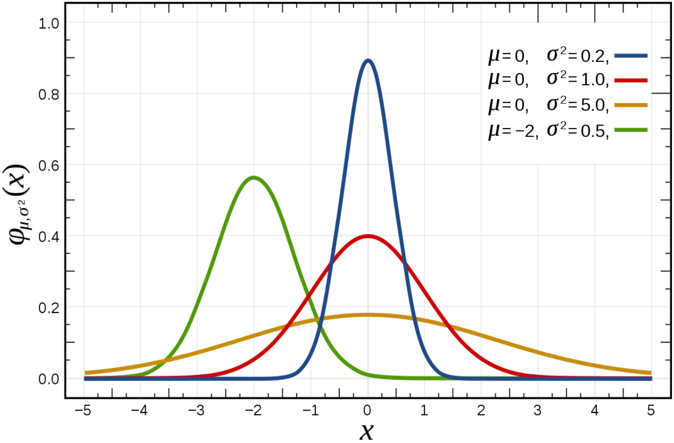

Figure 3: Densités de lois gaussiennes pour différents paramètres

Loi Gamma

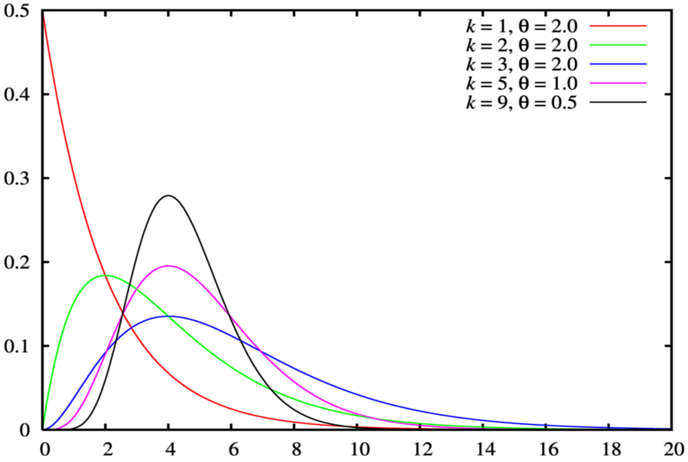

Figure 4: Densité de lois gamma pour différents paramètres

Fonction de répartition

Autre façon de décrire la loi de X est par sa fonction de répartition.

Fonction de répartition de la variable aléatoire X la fonction F définie sur $\mathbb{R}$

$$F(t) = P(X \leq t).$$

- F est une fonction croissante et définie sur R à valeurs dans $[0, 1]$.
- Si X **v.a. dicrète**: F est une fonction en escalier.
- Si X **v.a. continue**: F est une fonction continue et la dérivée de F est f la **densité de probabilité**.

2. Vecteurs aléatoires discrets

Couple de variables/vecteur aléatoire (X, Y)

- **Loi de probabilité jointe** de (X, Y)

$$\{P(X = x, Y = y)\}_{(x,y)\in Supp_X \times Supp_Y} .$$

- $P(X = x, Y = y) \in [0, 1]$.
- $\sum_{x\in supp_X} \sum_{y\in supp_Y} P(X = x, Y = y) = 1$.

Exemple: On lance 2 dés non truqués.

- $Y=$ *"le nombre de chiffres impairs apparus"*, $Supp_Y = \{0, 1, 2\}$.
- $X=$ *"la somme des 2 chiffres"*, $Supp_X = \{2, \cdots, 12\}$.

Lois marginales

Pour trouver la loi de X (et réciproquement de Y) à partir de la loi du couple (X, Y) il suffit de sommer sur toutes les valeurs de Y (et réciproquement de X)!!

Lois marginales

- Pour tout $x \in Suup_X$,

$$P(X = x) = \sum_{y \in Supp_Y} P(X = x, Y = y)$$

- Pour tout $y \in Suup_Y$,

$$P(Y = y) = \sum_{x \in Supp_X} P(X = x, Y = y)$$

Exemple de loi jointe: cas discret

Prenons un exemple, soit (X, Y) un couple de variables aléatoires discrètes dont la loi est résumée dans le tableau suivant :

$Y \mid X$	0	1	2
-1	0.2	0.1	0.1
1	0.1	0	0.5

$Y = y$	-1	1
$P(Y = y)$	0.4	0.6

$X = x$	0	1	2
$P(X = x)$	0.3	0.1	0.6

3. Indépendance

X et Y sont indépendantes si et seulement si

Pour tout $x \in Supp_X$ **et pour tout** $y \in Supp_Y$

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

4. Loi conditionnelle

Loi conditionnelle de Y sachant $\{X = x\}$:

$$\{P(Y = y|X = x)\}_{y \in Supp_Y}$$

- Pour tout $y \in Supp_Y$ et pour tout $X \in Supp_X$

$$P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} \text{ et } P(Y = y|X = x) \in [0, 1]$$

- $\sum_{y \in supp_Y} P(Y = y|X = x) = 1.$

Loi conditionnelle: Exemple

Loi jointe de (X, Y)

$Y \mid X$	0	1	2
-1	0.2	0.1	0.1
1	0.1	0	0.5

Loi (marginale) de X

$X = x$	0	1	2
$P(X = x)$	0.3	0.1	0.6

Loi conditionnelle de Y sachant $X = 0$:

- $P(Y = 1|X = 0) = \frac{P(Y=1,X=0)}{P(X=0)} = \frac{0.1}{0.3} = \frac{1}{3}$.
- $P(Y = -1|X = 0) = \frac{P(Y=-1,X=0)}{P(X=0)} = \frac{0.2}{0.3} = \frac{2}{3}$.

5. Covariance et coefficient de corrélation linéaire

Covariance de X et de Y

Soient X et Y deux v.a. de variance finie non nulle et d'espérance finie.

$$Cov(X,Y) = \sum_{x \in Supp_X} \sum_{y \in Supp_Y} (x - E[X])(y - E[Y]) P(X = x, Y = y).$$

• Symétrie: $Cov(X,Y) = Cov(Y,X)$
• $Cov(X,X) = V(X)$
• Si X et Y deux v.a. indépendantes $\Rightarrow Cov(X,Y) = 0$
(Réciproque fausse sauf pour vecteur gaussien)

Covariance et coefficient de corrélation linéaire

Coefficient de corrélation linéaire de X et de Y

Soient X et Y deux v.a. de variance finie non nulle et d'espérance finie.

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}.$$

• $|\rho(X, Y)| \leq 1$.
• Renseigne sur l'existence d'une éventuelle relation affine entre X et Y. ($|\rho(X, Y)| = 1 \Rightarrow$ lien linéaire)

6. Théorèmes de convergence

Loi des grands nombres

Soit une suite de v.a. $X_1, X_2, \cdots$ i.i.d. telles que $E[X_i] = m < \infty$ alors

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \xrightarrow[n\to+\infty]{P} m.$$

Théorème central limite (TCL)

Soit une suite de v.a. $X_1, X_2, \cdots$ i.i.d. telles que $E[X_i] = m < \infty$ et $V(X_i) = \sigma^2 < \infty$ alors

$$\sqrt{n}\frac{\overline{X} - m}{\sigma} \xrightarrow[n\to+\infty]{Loi} \mathcal{N}(0, 1).$$

7. Parlons R

Avantages de R/Rstudio

- Méthodes récentes
- Multi-plateforme
- Gratuit
- Installation : http ://www.r-project.org
- `install.packages("tidyverse")`

Quelques commandes élémentaires à connaître

Commandes élémentaires

`max()`	maximum
`min()`	minimum
`sum()`	somme
`cumsum()`	sommes cumulées
`mean()`	moyenne
`median()`	médiane
`sort()`	tri par ordre croissant
`sd()`	écart-type
`cor()`	corrélation
`which.max()`	qui atteint le max

Quelques lois élémentaires

Quelques lois usuelles

`unif(.,.)`	Uniforme (min, max)
`geom(.)`	Géométrique (proba)
`norm(.,.)`	Normale (moyenne, écart type)
`binom(.,.)`	Binomiale (taille, proba)
`exp(.)`	Exponentielle (taux)
`pois(.)`	Poisson (λ)
`beta(.,.)`	Beta (forme1, forme2)
`t(.)`	Student (degré de liberté)
`chisq(.)`	Chi2 (degré de liberté)
`f(.,.)`	Fisher (degré de liberté,degré de liberté)
`gamma(.,.)`	Gamma (forme/shape, echelle/rate)

8. Petites simulations sous R

Ouvrir Rstudio et ouvrir le fichier Participant1.Rmd

Fixer la *seed*

Question 1: Fixer la "graine" aléatoire (random number generator) ce qui de fixer les résultats.

```
set.seed(1234)
```

TP2

Question 2: **Tirage selon la loi "$*$": `r*`.** Tirer, par exemple, au hasard n=5 valeurs selon une loi normale $\mathcal{N}(m, \sigma)$ de moyenne mean=10 et d'écart type sd=2. (commande `rnorm(n,mean,sd)`).

TP2

Question 2: **Tirage selon la loi "$*$": `r*`.** Tirer, par exemple, au hasard n=5 valeurs selon une loi normale $\mathcal{N}(m,\sigma)$ de moyenne mean=10 et d'écart type sd=2. (commande `rnorm(n,mean,sd)`).

```
rnorm(5,10,2)
```

```
## [1]  7.585869 10.554858 12.168882  5.308605 10.858249
```

TP3

Question 3: **Quartiles de la loi "$*$": `q*`.** Afficher le premier quartile (ordre 0.25) de la loi $\mathcal{N}(mean, sd)$ de moyenne mean$=$10 et d'écart type sd$=$2. (commande `qnorm(ordre,mean,sd)`).

Question 3: **Quartiles de la loi "$*$": q$*$.** Afficher le premier quartile (ordre 0.25) de la loi $\mathcal{N}(mean, sd)$ de moyenne mean=10 et d'écart type sd=2. (commande `qnorm(ordre,mean,sd)`).

```
qnorm(0.25,10,2)
```

```
## [1] 8.65102
```

TP4

Question 4: **Probabilité ponctuelle de la loi "$*$": `d*`.** Afficher les probablités qu'une v.a. X de loi Poisson de paramètre $\lambda = 4$; $\mathcal{P}(\lambda)$)a de prendre les valeurs 0, 1 et 2. (commande `dpois(0:2,lambda)`.

TP4

Question 4: **Probabilité ponctuelle de la loi "$*$": `d*`.** Afficher les probablités qu'une v.a. X de loi Poisson de paramètre $\lambda = 4$ ($\mathcal{P}(\lambda)$) a de prendre les valeurs 0, 1 et 2. (commande `dpois(0:2,lambda)`.

```
dpois(0:2, 4)
```

```
## [1] 0.01831564 0.07326256 0.14652511
```

TP5

Question 5: **Evaluer plusieurs probabilité d'un coup :** Afficher les proba. d'une v.a. binomiale $\mathcal{B}(n, p)$, $n = 10$ et $p = 1/4$. (commande `dbinom(0:n, n,p)`). Vérifier que la somme fait 1.

Question 5: **Evaluer plusieurs probabilité d'un coup :** Afficher les proba. d'une v.a. binomiale $\mathcal{B}(n,p)$, $n = 10$ et $p = 1/4$. (commande `dbinom(0:n, n,p)`). Vérifier que la somme fait 1.

```
V=dbinom(0:10, 10, 0.25); V
```

```
##  [1] 5.631351e-02 1.877117e-01 2.815676e-01 2.502823e-01 1.459980e-01
##  [6] 5.839920e-02 1.622200e-02 3.089905e-03 3.862381e-04 2.861023e-05
## [11] 9.536743e-07
```

```
sum(V)
```

```
## [1] 1
```

TP6

Question 6: **Valeur de la fonction de répartition de la loi"$*$": `d*`.** Afficher la probablité qu'une v.a. X de loi Normale $\mathcal{N}(mean, sd)$ de moyenne mean=10 et d'écart type sd=2 soit inférieure à 12. (commande `pnorm(12,mean,sd)`).

TP6

Question 6: **Valeur de la fonction de répartition de la loi"$*$": `d`$*$.** Afficher la probablité qu'une v.a. X de loi Normale $\mathcal{N}(mean, sd)$ de moyenne mean=10 et d'écart type sd=2 soit inférieure à 12. (commande `pnorm(12,mean,sd)`).

```
pnorm(12,10,2)
```

```
## [1] 0.8413447
```