Variables aléatoires: Généralités

LMAFY1101

Anouar El Ghouch

LSBA, UCLouvain

Plan

Variable discrète et variable continue

La distribution d'une variable aléatoire

Espérance, variance et quantiles

Variable discrète et variable continue

Il est souvent plus pratique et plus intéressant d'associer une valeur numérique au résultat d'une expérience aléatoire, plutôt que de travailler directement avec des événements particuliers.

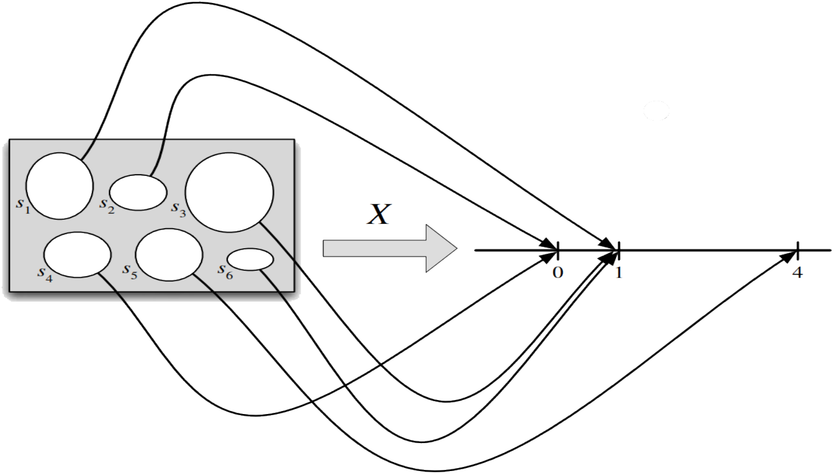
Exemple. Si on lance une pièce de monnaie trois fois, l'univers est :

$$S = \{\mathrm{PPP}, \mathrm{PPF}, \mathrm{PFP}, \mathrm{FPP}, \mathrm{PFF}, \mathrm{FPF}, \mathrm{FFP}, \mathrm{FFF}\}$$

À partir de cet ensemble et en fonction de notre intérêt, nous pouvons définir diverses variables aléatoires :

- $X =$ nombre total de piles;
- $Y =$ nombre de piles lors des deux premiers essais ;
- $Z = 1$ si le premier lancer est pile et 0 sinon.

En langage mathématique une variable aléatoire (**v.a.**) est une fonction de l'ensemble S des résultats possibles d'une expérience aléatoire dans l'ensemble des réels $\mathbb{R}$:



Dans l'exemple avec la pièce de monnaie lancée trois fois :

$$
\begin{aligned}
X :\ S &\rightarrow \{0, 1, 2, 3\} \\
w &\rightarrow \mapsto X(w),
\end{aligned}
$$

Avec

- $X(PPP) = 3$
- $X(PPF) = X(PFP) = X(FPP) = 2$
- $X(PFF) = X(FPF) = X(FFP) = 1$
- $X(FFF) = 0$

Autrement dit,

- $X = 3 \Leftrightarrow \{PPP\}$
- $X = 2 \Leftrightarrow \{PPF, PFP, FPP\}$
- $X = 1 \Leftrightarrow \{PFF, FPF, FFP\}$
- $X = 0 \Leftrightarrow \{FFF\}$

L'ensemble des valeurs possibles d'une v.a. constitue son domaine ou univers.

Dans notre exemple, le domaine de X est $\{0, 1, 2, 3\}$.

V.A. DISCRÈTE

Si cet ensemble est dénombrable (fini ou infini), on dit que la v.a. est **discrète**.

V.A. CONTINUE

Si cet ensemble est un intervalle de $\mathbb{R}$, on dit que la v.a. est **continue**. Une v.a. continue peut prendre toutes les valeurs réelles possibles entre un minimum et un maximum.

EXEMPLES

- X = « le nombre de lancers nécessaires pour obtenir un premier SIX dans une séquence de lancers d'un dé bien balancé » $\in \{1, 2, 3, \ldots\} \Rightarrow X$ est discrète.
- L'âge d'un individu pris au hasard $\in [0\ \mathrm{ans}, 100\ \mathrm{ans}] \implies$ c'est une v.a. continue.
- La quantité de précipitation (en mm) en juillet prochain en Belgique $\in [0\ \mathrm{mm}, 1000\ \mathrm{mm}] \implies$ c'est une v.a. continue.

La distribution d'une variable aléatoire

Définition. La fonction de distribution de probabilité de la **variable discrète** X est la fonction $p_X(x) \equiv p(x)$ qui associe à chaque valeur possible x de X sa probabilité $P(X = x)$.

$$p_X(x) = P(X = x) = \text{ la probabilité que la v.a. } X \text{ prenne la valeur particulière } x.$$

Cette fonction satisfait toujours les deux conditions suivantes :

$$p(x) \geqslant 0 \text{ et } \sum_x p(x) = 1.$$

Si I est un ensemble quelconque de nombres réels, alors la probabilité que la v.a. X prenne une valeur dans I est donnée par l'équation

$$P(X \in I) = \sum_{x \in I} p(x)$$

EXEMPLE. On lance deux pièces de monnaie. Soit $X =$ nombre total de piles.

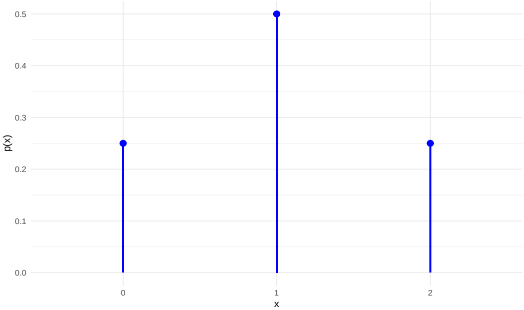
La distribution de X peut être résumée sous forme d'un

- Tableau

x	0	1	2
$P(X = x)$	$1/4$	$1/2$	$1/4$

- Graphique

```
> x <- c(0, 1, 2); y <- c(1 / 4, 1 / 2, 1 / 4)
> plot(y ~ x, type = "h"); points(x, y)
```



Exercice

Calculez la distribution de la v.a. $Y = (X - 1)^2$

Exercice

Calculez la distribution de la v.a. $Y = (X-1)^2$

Réponse

y	0	1
$P(Y=y)$	$1/2$	$1/2$

La fonction de distribution cumulative (CDF ou fonction de répartition) d'une variable aléatoire X (discrète ou continue) indique, pour chaque x, la probabilité que X prenne une valeur au plus égale à x.

$$\begin{aligned} F_X : \mathbb{R} &\to [0, 1] \\ x &\to P(X \leqslant x) \end{aligned}$$

Notez qu'il s'agit d'une fonction croissante à valeurs dans $[0, 1]$.

Dans le cas discret, si $X \in \{x_1, x_2, \ldots\}$ avec $x_1 \leqslant x_2 \leqslant \ldots$, alors

- $$F_X(x) = \sum_{i : x_i \leqslant x} P(X = x_i)$$
- $$P(X = x_i) = F_X(x_i) - F_X(x_{i-1})$$

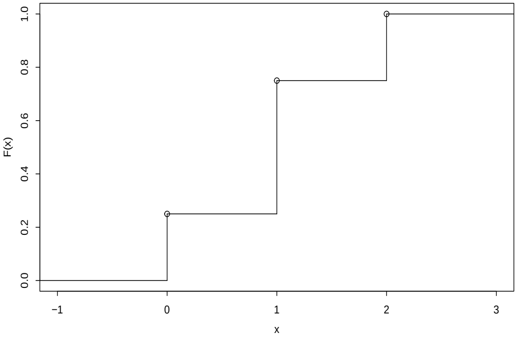
Exemple

On lance deux pièces de monnaie. Soit $X =$ nombre total de piles.

x	0	1	2
$P(X = x)$	$1/4$	$1/2$	$1/4$
$P(X \leqslant x)$	$1/4$	$3/4$	1

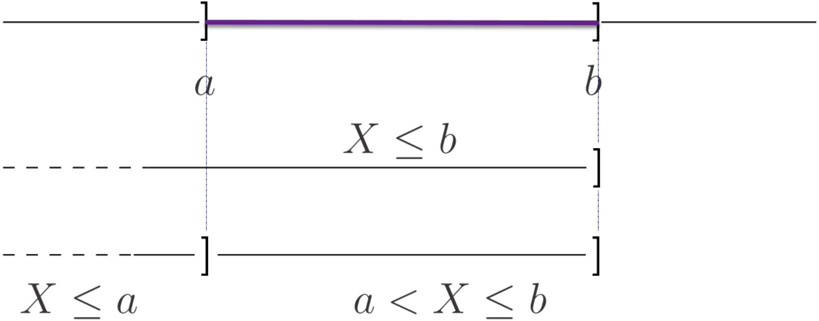
$$F(x) = \begin{cases} 0\,, & \text{si } x < 0 \\ 0.25\,, & \text{si } 0 \leqslant x < 1 \\ 0.75\,, & \text{si } 1 \leqslant x < 2 \\ 1\,, & \text{si } x \geqslant 2 \end{cases}$$

```
> F <- stepfun(c(0, 1, 2), c(0, 1/4, 3/4, 1))
> plot(F, main = "", xlab = "x", ylab = "F(x)")
```



Quelle que soit la v.a. X (discrète ou continue) et pour tout a et b tels que $a \leqslant b$, on a

$$P(a < X \leqslant b) = F(b) - F(a)$$



Voici la CDF de $X \in \{2, \ldots, 12\}$

x	2	3	4	5	6	7	8	9	10	11	12
$F(x)$	1/36	2/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	36/36

Calculez $P(X = 5)$ et $P(3 \leqslant X < 7)$?

Exercice

Voici la CDF de $X \in \{2, \ldots, 12\}$

x	2	3	4	5	6	7	8	9	10	11	12
$F(x)$	1/36	2/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	36/36

Calculez $P(X = 5)$ et $P(3 \leqslant X < 7)$?

Solution:

$$P(X = 5) = F(5) - F(4) = 4/36$$
$$P(3 \leqslant X < 7) = F(6) - F(2) = 14/36.$$

On désire modéliser l'expérience suivante : ≪ Lancer une aiguille au hasard dans le segment $[0, 1]$≫. On suppose qu'une fois lancée, l'aiguille peut s'arrêter n'importe où dans $[0, 1]$, avec la même probabilité. Soit X la position exacte du point choisit.

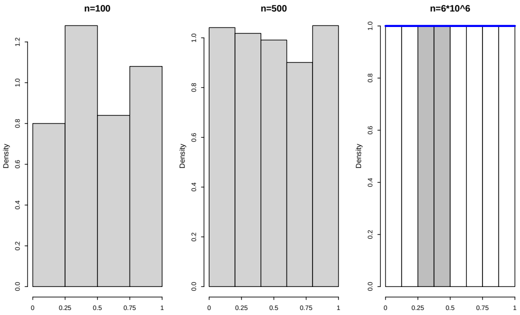
Il est raisonnable de supposer que la probabilité que l'aiguille se trouve dans l'intervalle $[a, b]$ $(0 \leqslant a < b \leqslant 1)$ ne dépend que de sa longueur :

$$P(a \leqslant X \leqslant b) = b - a$$

Dans ce cas, $P(X = x) = \lim_{\epsilon \to 0^+} P(x - \epsilon \leqslant X \leqslant x + \epsilon) = 0$.

Pour toute v.a. continue X, $\boxed{P(X = x) = 0}$, $\forall x$.

À l'aide de R, on a simulé cette expérience 100, 500 et 6×10^6 fois. Voici l'histogramme des résultats.



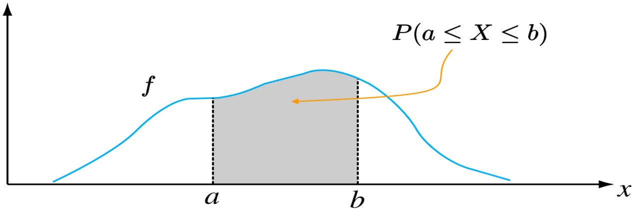
$$P(0.25 \leqslant X \leqslant 0.5) = 0.25 = \int_{0.25}^{0.5} f(x)dx,$$

où

$$f(x) = I(0 \leqslant x \leqslant 1) = \begin{cases} 1\,, & \text{si } 0 \leqslant x \leqslant 1 \\ 0\,, & \text{sinon} \end{cases}$$

Définition. La fonction de densité d'une v.a. **continue** X est la fonction $f_X \equiv f$ telle que pour tout $a, b \in \mathbb{R}$

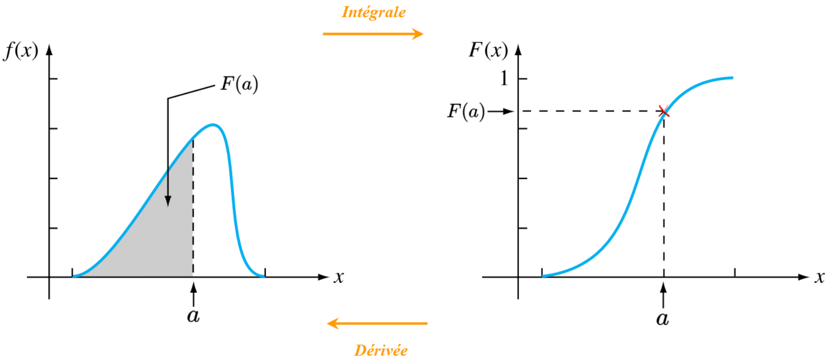
$$\boxed{P(a \leqslant X \leqslant b) = \int_a^b f_X(x)dx}$$



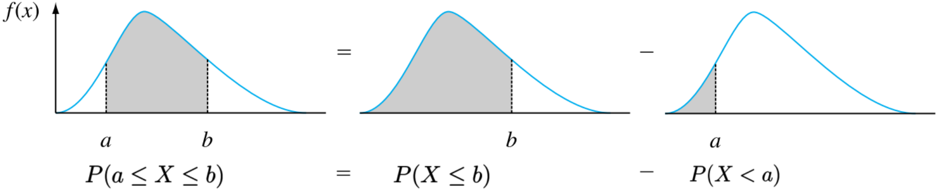
Quelques propriétés

- $f(x) \geqslant 0$
- $\displaystyle\int_{-\infty}^{+\infty} f(x)dx = 1$

Relation entre f et F



- $F(a) = P(X \leqslant a) = P(X < a) = \int_{-\infty}^{a} f(t)dt$
- $f(a) = F^{'}(a)$



$$P(a \leqslant X \leqslant b) = F(b) - F(a) = \int_a^b f(x)dx$$

Exemple 1

Supposons que la durée de vie (en heures) d'une ampoule électrique soit un phénomène aléatoire représenté par la densité de probabilité:

$$f(x) = \begin{cases} \frac{1}{1000} e^{-x/1000}, & \text{si } x \geqslant 0 \\ 0, & \text{si } x < 0 \end{cases}$$

Calculons (a) la probabilité que la durée de vie soit comprise entre 100 et 1000 heures et (b) la probabilité d'avoir une durée de vie supérieure à 1000 heures.

Exemple 1

Supposons que la durée de vie (en heures) d'une ampoule électrique soit un phénomène aléatoire représenté par la densité de probabilité:

$$f(x) = \begin{cases} \dfrac{1}{1000} e^{-x/1000}, & \text{si } x \geqslant 0 \\ 0, & \text{si } x < 0 \end{cases}$$

Calculons (a) la probabilité que la durée de vie soit comprise entre 100 et 1000 heures et (b) la probabilité d'avoir une durée de vie supérieure à 1000 heures.

$$\begin{aligned} \text{(a) } P(100 \leqslant X \leqslant 1000) &= \int_{100}^{1000} \frac{1}{1000} e^{-x/1000} dx \\ &= -e^{-x/1000}]_{100}^{1000} = -e^{-1} + e^{-0.1} = 0.537 \\ \text{(b) } P(X > 1000) &= \int_{1000}^{\infty} \frac{1}{1000} e^{-x/1000} dx \\ &= -e^{-x/1000}]_{1000}^{\infty} = 0 + e^{-1} = 0.368. \end{aligned}$$

Vous pouvez aussi utiliser R pour calculer les intégrales. Mais vous devez faire attention, puisqu'il s'agit d'approximation numérique et non pas de valeurs exactes. Voici un exemple.

```
> Fun <- function(x) (1/1000) * exp(-x/1000)

> integrate(Fun, lower = 100, upper = 1000)

0.537 with absolute error < 6e-15

> integrate(Fun, lower = 1000, upper = Inf)

0.368 with absolute error < 4.1e-06
```

Exemple 2

On choisit un point au hasard et de façon uniforme sur un disque de rayon égal à cinq centimètres et on pose $X = \ll$la distance entre le point choisi et le centre du disque$\gg \in [0, 5]$

Calculons la fonction de répartition et la fonction de densité de X.

Exemple 2

On choisit un point au hasard et de façon uniforme sur un disque de rayon égal à cinq centimètres et on pose X = «la distance entre le point choisi et le centre du disque» $\in [0, 5]$

Calculons la fonction de répartition et la fonction de densité de X.

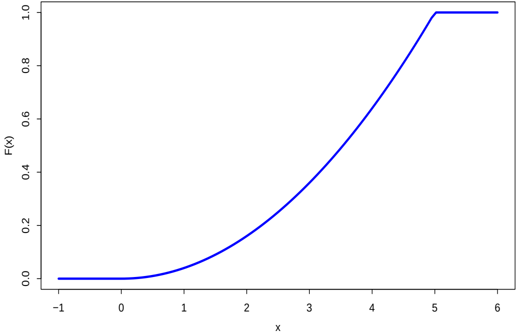
Clairement, pour $0 \leqslant x \leqslant 5$ on a que $F(x) = \dfrac{\pi \times x^2}{\pi \times 5^2} = \dfrac{x^2}{25}$.

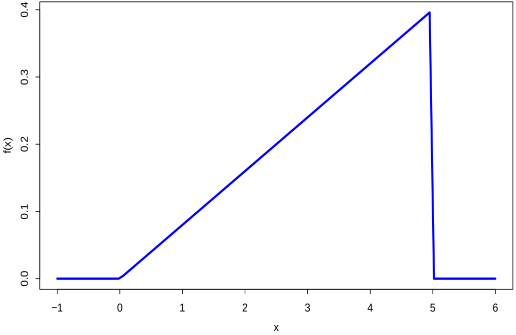
$$\Rightarrow F(x) = \begin{cases} 0\,, & \text{si } x < 0 \\ x^2/25\,, & \text{si } 0 \leqslant x \leqslant 5 \\ 1\,, & \text{si } x > 5 \end{cases}$$

En dérivant, on obtient la densité

$$f(x) = \frac{2x}{25} I(0 \leqslant x \leqslant 5)$$

```
> F <- function(x) (x^2/25) * ((0 <= x) * (x <= 5)) + (x > 5)
> curve(F, from = -1, to = 6)
> f <- function(x) (x/12.5) * (x >= 0) * (x <= 5)
> curve(f, from = -1, to = 6)
```





$$P(2 \leqslant X \leqslant 4) = F(4) - F(2) = 16/25 - 4/25 = 12/25$$

$$= \int_2^4 f(x)\,dx = \int_2^4 \frac{2x}{25}\,dx = 12/25$$

Exercice

Calculez la densité de la v.a. $Y = \log(X)$

Exercice

Calculez la densité de la v.a. $Y = \log(X)$

Réponse

(1) On commence par calculer $F_Y(y) = P(Y \leqslant y)$

$$F_Y(y) = P(X \leqslant \exp(y)) = \begin{cases} \exp(2y)/25\,, & \text{si } y \leqslant \log(5) \\ 1\,, & \text{si } y > \log(5) \end{cases}$$

(2) On dérive pour obtenir f_Y

$$f_Y(y) = \frac{2}{25}\exp(2y)I(y \leqslant \log(5))$$

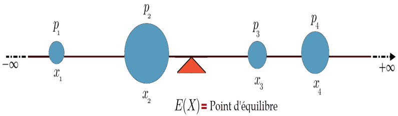
Espérance, variance et quantiles

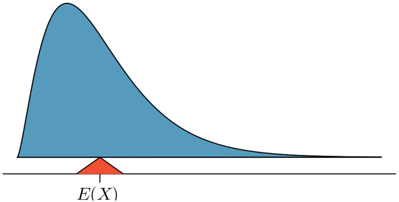
Exemple. Une urne contient 100 boules : 98 blanches et 2 noires. On tire au hasard une boule. On gagne 100 euros si elle est noire et 50 euros si elle est blanche. A quel gain devons-nous nous attendre ?

- Pourquoi pas $(100 + 50)/2$?
- Les chances de gagner 100 euros sont beaucoup plus faibles que celles de gagner 50 euros.
- On gagne 100 euros avec la probabilité de $2/100$ et 50 euros avec la probabilité de $98/100$.
- Le gain espéré est la moyenne des gains (100 et 50) pondérés par les probabilités respectives (0.02 et 0.98):

$$100 \times 0.02 + 50 \times 0.98 = 51.$$

Pour une v.a. X, on appelle l'espérance de X, notée par $\mathbb{E}(X)$ ou μ_X, la quantité $\boxed{\mu_X = \sum_x x\, p(x)}$, si X est discrète ou $\boxed{\mu_X = \int_{-\infty}^{+\infty} x\, f(x) dx}$, si X est continue





Dans l'exemple des deux pièces

$$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1$$

```
sum(c(0, 1, 2) * c(1 / 4, 1 / 2, 1 / 4))
```

Dans l'exemple du disque

$$E(X) = \int_0^5 x \frac{2x}{25} dx = 10/3$$

```
integrate(function(x) x * (2 * x) / 25, 0, 5)
```

Interprétation de l'espérance: lois des grands nombres

On peut interpréter l'espérance d'une variable aléatoire comme si c'était la moyenne (empirique) qu'on obtiendrait si on répétait notre expérience aléatoire un très très grand nombre de fois.

Ainsi, si $x_1, x_2, \ldots, x_n$ sont n réalisations indépendantes de la v.a. X, alors

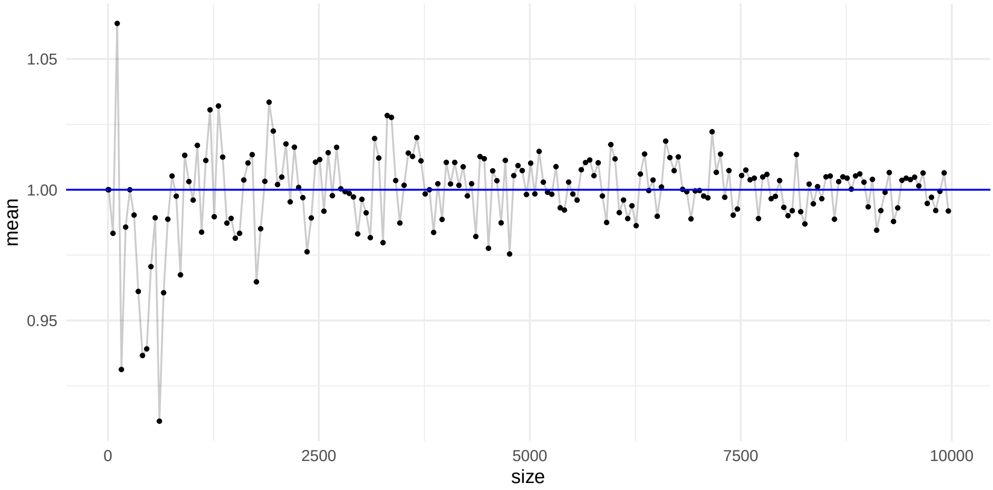
$$E(X) = \lim_{n \to \infty} \bar{x}_n.$$

À l'aide de R, on peut vérifier cela facilement pour, par exemple, notre expérience avec les deux pièces, où $X =$ Nombre total de piles.

```
> sample(c(0, 1, 2), size = 50000, replace = TRUE, prob = c(1 / 4, 1 / 2, 1 / 4)) |>
    mean()
```

```
[1] 0.999
```

Lorsque on tourne le code précédent avec $size = 2, 10, 60, 110, \ldots, 9960$, on obtient les résultats présentés dans la figure suivante.



Espérance et transformation

On considère une v.a. X et une fonction g. On aimerait connaitre l'espérance de la nouvelle variable $g(X)$. On a la proposition suivante

$$E(g(X)) = \sum_x g(x)p(x)\,, \quad \text{si } X \text{ discrète}$$

$$E(g(X)) = \int g(x)f(x)dx\,, \quad \text{si } X \text{ continue}$$

Dans le cas de l'exemple avec les deux pièces de monnaie, calculons $E(Y)$ pour $Y = (X-1)^2$, à l'aide de (i) la définition et (ii) la proposition ci-dessus

(i) $E(Y) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = 0.5$ voir ceci

(ii) $E(X-1)^2 = (0-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{1}{2} + (2-1)^2 \times \frac{1}{4} = 0.5$

Dans le cas de l'exemple du disque, calculons $E(Y)$ pour $Y = \log(X)$

(i) $E(Y) = \frac{2}{25}\int_{-\infty}^{\log(5)} y\exp(2y)dy \approx 1.1$ voir ceci

(ii) $E(\log(X)) = \frac{2}{25}\int_0^5 \log(x)xdx \approx 1.1$

Pour une v.a. X, on appelle l'variance de X, notée par $\mathbb{V}ar(X)$ ou σ_X^2, la quantité

$$\sigma_X^2 = E(X - \mu_X)^2$$

σ_X, la racine carrée de la variance est appelée l'**écart-type** de la v. a. X.

- Si X est discrète

$$\sigma_X^2 = \sum_x (x - \mu_X)^2 \; p(x)$$

Dans l'exemple des deux pièces:

$$\sigma_X^2 = E(X - 1)^2 = 0.5$$

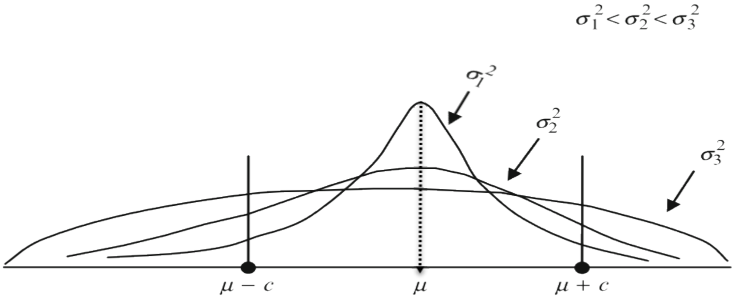
- Si X est continue

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 \; f(x) dx$$

Dans l'exemple du disque :

$$\sigma_X^2 = \int_0^5 (x - 10/3)^2 \frac{2x}{25} = 25/28$$

Les densités avec une variance/un écart-type plus élevée sont plus étalées que les densités avec une variance plus faible. Ce qui se traduit en pratique par des données plus dispersées.



Voici une inégalité mathématique appelé *l'inégalité Chebyshev* qui permet de mieux comprendre la variance. Pour toute v.a. X de moyenne (espérance) μ, et toute constante $c > 0$, on a que

$$\boxed{P(|X - \mu| \leqslant c) \geqslant 1 - \sigma^2/c^2}$$

Ainsi, $P(\mu - c \leqslant X \leqslant \mu + c)$ est d'au moins

$$\begin{aligned} 75\% &\text{ pour } c = 2\sigma \\ 89\% &\text{ pour } c = 3\sigma \\ 94\% &\text{ pour } c = 4\sigma \end{aligned}$$

Aussi, comme pour la moyenne, si $x_1, x_2, \ldots, x_n$ sont n réalisations indépendantes de la v.a. X, alors

$$\sigma_X^2 = \lim_{n \to \infty} s_n^2.$$

On peut donc interpréter la variance (l'écart-type) d'une variable aléatoire comme si c'était la variance (écart-type) empirique qu'on obtiendrait si on répétait notre expérience aléatoire un très très grand nombre de fois.

À l'aide de R, on peut vérifier cela facilement pour, par exemple, notre expérience avec les deux pièces, où $X =$ Nombre total de piles.

```
> sample(c(0, 1, 2), size = 50000, replace = TRUE, prob = c(1 / 4, 1 / 2, 1 / 4)) |>
    var()

[1] 0.5
```

Voici quelques propriétés de l'espérance et de la variance

- $E(a + bX) = a + bE(X)$, mais de façon générale $E(g(X)) \neq g(E(X))$.
- $E(X + Y) = E(X) + E(Y)$, quelles que soient les v.a. X et Y.
- $Var(X) = E(X^2) - (E(X))^2$.
- $Var(a + bX) = b^2 Var(X)$.
- $Var(X + Y) = Var(X) + Var(Y)$, si les v.a. X et Y sont indépendantes.

Deux v.a. X et Y sont indépendantes si le fait de savoir qu'un événement impliquant l'une a eu lieu n'apporte aucune information sur l'occurrence de tout autre événement impliquant l'autre. On peut exprimer cela mathématiquement par la relation:

$$P(a \leqslant Y \leqslant b | c \leqslant X \leqslant d) = P(a \leqslant Y \leqslant b),\ \forall a, b, c.$$

Si X et Y ne sont pas indépendantes, alors $Var(X + Y) \neq Var(X) + Var(Y)$.

Exemple. On sait que $Var(X + X) = Var(2X) = 4Var(X)$. Or, si l'on applique à tort la formule d'indépendance, on obtiendrait que $Var(X) + Var(X) = 2Var(X)$.

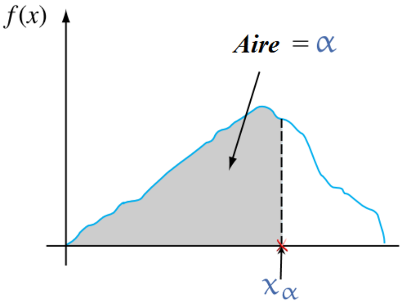
Quantiles d'une distribution

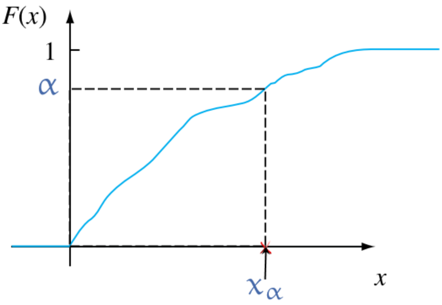
Soit α un chiffre donné compris entre 0 et 1, et soit X une v.a. continue. La quantité $q_X(\alpha) \equiv x_\alpha$ tel que

$$P(X \leqslant x_\alpha) = \alpha \Leftrightarrow x_\alpha = F_X^{-1}(\alpha)$$

est appelée le quantile d'ordre α de X.

x_α partage la population en deux groupes, $100 \times \alpha\%$ ont des valeurs en dessous et $100 \times (1-\alpha)\%$ ont des valeurs au-dessus.

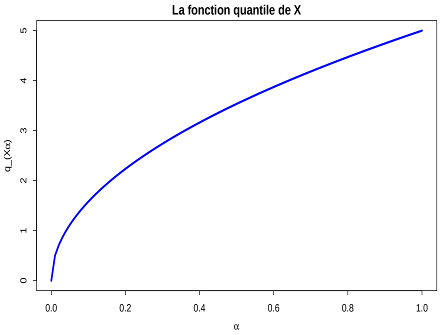




Dans l'exemple du disque avec $X =$ "la distance entre le point choisi et le centre du disque".

$$F(x) = \begin{cases} 0\,, & \text{si } x < 0 \\ x^2/25\,, & \text{si } 0 \leqslant x < 5 \\ 1\,, & \text{si } x > 5 \end{cases}$$

$$\begin{aligned} & F(x) = \alpha \\ & \Leftrightarrow \frac{x^2}{25} = \alpha,\ 0 \leqslant x \leqslant 5 \\ & \Leftrightarrow x = 5 \times \sqrt{\alpha} \\ & \Rightarrow q_X(\alpha) = 5\sqrt{\alpha},\ \forall \alpha \in [0, 1]. \end{aligned}$$



Commandes R les plus utiles vues dans ce chapitre

- plot, points, curve
- stepfun, function
- integrate
- sum, mean, var
- sample