

§ 3 Distributions théoriques à une variable discrète

Packages de l'auteur

- On peut consulter le mode d'emploi du package **Statistique**:
<https://www.deleze.name/marcel/sec2/applmaths/packages/aide/Statistique.pdf>
- Avant d'utiliser le package, il faut le charger en donnant son adresse web:

```
Needs ["Statistique`",  
|nécessite  
"https://www.deleze.name/marcel/sec2/applmaths/packages/Statistique.m"]
```

Voici la liste des instructions disponibles :

```
Names ["Statistique`*"]  
|noms
```

```
{amplitudes, densiteContinue, densites, diagrammeBatons,  
diagrammeCumulatif, distributionContinue, distributionLissee, fctDensite,  
fctFrequenceCumulee, frequenceCumuleeContinue, frequenceCumuleeLissee,  
histogramme, InterpolatedQuantile, noeudsPolygonaux, polygoneDeDensite,  
quantileC, quantileLisse, sommesCumulees, StandardDeviationMLE, VarianceMLE}
```

- Le package **Tableaux** contient des commandes qui facilitent la présentation des données et résultats sous la forme de tableaux:

```
Needs ["Tableaux`",  
|nécessite  
"https://www.deleze.name/marcel/sec2/applmaths/packages/Tableaux.m"]
```

```
Names ["Tableaux`*"]  
|noms
```

```
{afficheTableau, afficheTableauTitre, arrondis, fusionneColonnes,  
fusionneLignes, fusionneTableaux, prodCart, prodCartTrans, tableauGraph}
```

- On peut consulter le mode d'emploi du package **Tableaux**:
<https://www.deleze.name/marcel/sec2/applmaths/packages/aide/Tableaux.pdf>

Pour ne pas oublier d'exécuter ces instructions au début de chaque session de travail, il est conseillé de déclarer les instructions **Needs** comme étant des cellules d'initialisation. Pour ce faire, sélectionnez les cellules voulues puis passez par le menu

Cell / Cell properties / Initialization cell

§ 3.1 Distribution de Bernoulli

Expérience aléatoire

Lançons un dé et notons *le nombre de six*. Plus précisément, notons

1 (succès) si le dé montre six;

0 (échec) dans les autres cas.

On dit que la distribution de cette variable statistique est une *distribution de Bernoulli*. En faisant l'expérience, on pourrait obtenir un échantillon.

Simulation et nombres pseudo-aléatoires

Les systèmes informatiques sont capables de générer des nombres pseudo-aléatoires qui sont semblables à ceux produits par l'expérience. Voici par exemple comment produire un échantillon de taille n :

$n = 40$;

? BernoulliDistribution

BernoulliDistribution[p] represents a Bernoulli distribution with probability parameter p . >>

```
bern = BernoulliDistribution[ $\frac{1}{6}$ ];
```

[distribution de Bernoulli]

```
x = RandomInteger[bern, n]
```

[entier aléatoire]

```
{0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0}
```

Comme avec le dé, une nouvelle exécution donnera probablement un résultat différent:

```
x = RandomInteger[bern, n]
```

[entier aléatoire]

```
{0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0}
```

Probabilités

Du point de vue théorique, il y a deux modalités

$$c_0 = 0 \qquad c_1 = 1$$

dont probabilités sont respectivement

$$p_0 = \frac{5}{6} \qquad p_1 = \frac{1}{6}$$

Les probabilités représentent les fréquences théoriques. Pour la distribution de Bernoulli, on note plus couramment

$$p = p_1 = \text{probabilité de succès}$$

$$q = 1 - p = p_0 = \text{probabilité d'échec}$$

```
p =  $\frac{1}{6}$ ; N[p]
```

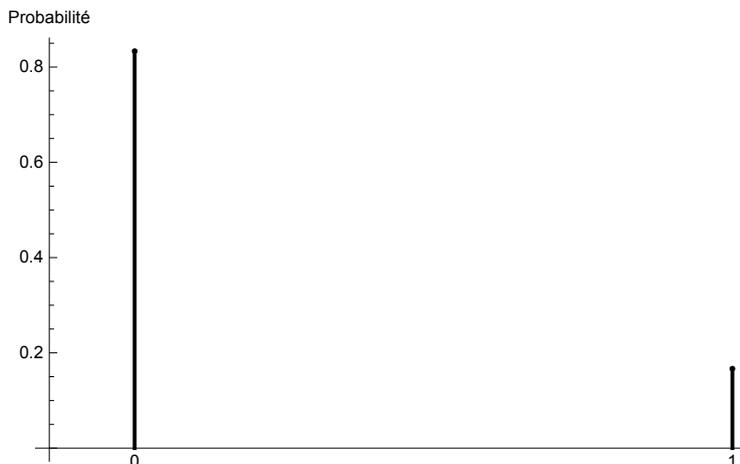
[valeur]

0.166667

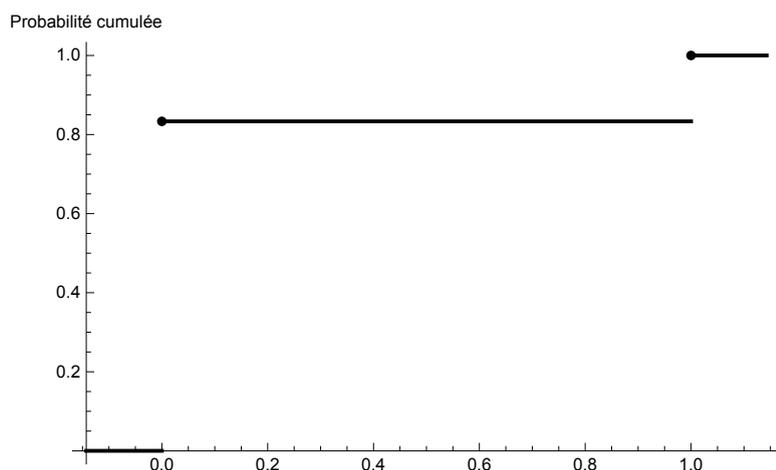
Modèle théorique

Le modèle théorique est la distribution construite avec les probabilités.

```
diagrammeBaton[{0, 1}, {5/6, 1/6}, AxesLabel -> {None, "Probabilité"}]
```



```
diagrammeCumulatif[{0, 1}, {5/6, 1/6}, AxesLabel -> {None, "Probabilité cumulée"}]
```



Rappelons que, à partir d'un échantillon, on peut calculer la moyenne empirique et la variance empirique avec les formules

$$m = c_0 f_0 + c_1 f_1 = 0 f_0 + 1 f_1 = f_1 = f = \text{fréquence de succès}$$

$$s^2 = (c_0 - m)^2 f_0 + (c_1 - m)^2 f_1 =$$

$$(0 - f_1)^2 f_0 + (1 - f_1)^2 f_1 = f^2 (1 - f) + (1 - f)^2 f = f (1 - f) (f + 1 - f) = f (1 - f)$$

Pour le modèle théorique, on remplace les fréquences empiriques par les probabilités.

La moyenne théorique est appelée **espérance mathématique de la variable aléatoire X**

$$E(X) = c_0 p_0 + c_1 p_1 = p = \text{probabilité de succès}$$

Pour définir la **variance théorique de la variable aléatoire X**, on remplace les fréquences par les probabilités et la moyenne empirique par l'espérance mathématique

$$V(X) = (c_0 - E(X))^2 p_0 + (c_1 - E(X))^2 p_1 = p(1 - p) = pq$$

Loi des grands nombres de Bernoulli

Pour un échantillon de taille n , la fréquence de succès est

$$f = N\left[\frac{1}{n} \text{Count}[\text{RandomInteger}[\text{bern}, n], 1]\right]$$

0.125

Comment les résultats dépendent-ils du nombre d'expériences réalisées ? Organisons une série de simulations

- une 1-ère simulation de 1 expérience,
- une 2-ème simulation de 10 expériences,
- une 3-ème simulation de 100 expériences,
- une 4-ème simulation de 1'000 expériences,
- une 5-ème simulation de 10'000 expériences,
- une 6-ème simulation de 100'000 expériences,
- une 7-ème simulation de 1'000'000 expériences,

et comparons chaque fois la fréquence empirique à la probabilité

$$\text{tabelle} = \text{Table}\left[\left\{\left\{10^t, N\left[\frac{1}{10^t} \text{Count}[\text{RandomInteger}[\text{bern}, 10^t], 1]\right], p\right\}, \{t, 0, 6\}\right\}\right]$$

```
{ {1, 0., 0.166667}, {10, 0., 0.166667}, {100, 0.15, 0.166667}, {1000, 0.175, 0.166667},
  {10000, 0.1606, 0.166667}, {100000, 0.16559, 0.166667}, {1000000, 0.166595, 0.166667} }
```

$$\text{afficheTableau}[\text{None}, \{\text{"n"}, \text{"Fréquences"}, \text{"Probabilité"}\}, \text{tabelle}]$$

<i>n</i>	<i>Fréquences</i>	<i>Probabilité</i>
1	0.	0.166667
10	0.	0.166667
100	0.15	0.166667
1000	0.175	0.166667
10 000	0.1606	0.166667
100 000	0.16559	0.166667
1 000 000	0.166595	0.166667

On peut observer que, lorsque le nombre d'expériences est faible, il peut y avoir d'importantes différences entre les fréquences empiriques et les probabilités. Par contre, quand le nombre d'expériences est élevé, les fréquences empiriques sont assez proches des probabilités. La loi des grands nombres de Bernoulli affirme que

Dans une série d'épreuves indépendantes, si le nombre d'expériences n tend vers l'infini, alors la suite des fréquences empiriques f_n converge en probabilité vers la probabilité p .

L'expression "converge en probabilité" signifie que, pour n'importe quel petit nombre positif ϵ (par exemple $\epsilon = 0.01$), la probabilité que l'erreur dépasse ϵ

$$P\{ |f_n - p| > \epsilon \}$$

tend vers 0 lorsque n tend vers l'infini.

La loi des grands nombres justifie la pratique qui consiste, lorsque la probabilité est inconnue et que la taille de l'échantillon est assez grande, d'utiliser la fréquence empirique comme estimateur de la probabilité

$$\hat{p} = f_n$$

Exercice 3.1 - 1 [Sans ordinateur]

Démontrez que, si une variable statistique ne peut prendre que les valeurs 0 ou 1, alors la moyenne empirique du carré est égale à la moyenne

$$m(x^2) = m(x)$$

Démontrez que, si une variable statistique ne peut prendre que les valeurs 0 ou 1, alors le calcul de la variance empirique peut se réduire à

$$s^2 = m(1 - m)$$

Démontrez les propriétés correspondantes de l'espérance mathématique et de la variance théorique

$$\begin{array}{lcl} X \text{ de Bernoulli} & \implies & E(X^2) = E(X) \\ X \text{ de Bernoulli} & \implies & V(X) = E(X)(1 - E(X)) \\ c' \text{ est } - \text{ à } - \text{ dire} & & \sigma^2 = \mu(1 - \mu) \end{array}$$

Exercice 3.1 - 2 [Sans ordinateur]

On a lancé 1000 fois une pièce de monnaie et on a obtenu 515 *face*.

Calculez la moyenne et l'écart-type de la variable statistique "nombre de *face* en un jet".

Exercice 3.1 - 3

Pour une variable X uniformément distribuée sur l'intervalle $[0; 1[$, on considère l'événement $0.4 \leq X < 0.6$.

- [Avec Mathematica] Pour un échantillon de taille n , calculez la fréquence m de l'événement ainsi que l'écart-type empirique s .
- [Sans ordinateur] Déterminez la probabilité de l'événement et l'écart-type théorique.

§ 3.2 Distribution binomiale

Distribution binomiale

La situation de notre exemple consiste en ceci

On lance 3 fois un dé et on compte le nombre de six obtenus.

Pour ce problème, nous connaissons la distribution théorique et nous pourrions comparer les paramètres empiriques et théoriques.

Le nombre de six obtenu avec **un dé** est

1 avec une probabilité de $\frac{1}{6}$

0 avec une probabilité de $\frac{5}{6}$

La distribution qui lui correspond est appelée *distribution de Bernoulli*.

La caractéristique des distributions de Bernoulli est de ne prendre que deux valeurs : 0 ou 1.

Les distributions de Bernoulli sont d'un usage fréquent avec le code

"1 = succès" avec la probabilité p

"0 = échec" avec la probabilité $q = 1 - p$.

Pour **3 dés**, il faut sommer 3 variables aléatoires de Bernoulli.

Chaque essai ayant une probabilité $p = \frac{1}{6}$ de succès (et $q = \frac{5}{6}$ d'échec), on réalise 3 essais indépendants et on compte le nombre de succès. La distribution théorique est appelée *distribution binomi-*

ale $(3, \frac{1}{6})$.

Plus généralement, on appelle *distribution binomiale* (k, p) la somme de k variables aléatoires indépendantes de Bernoulli de même probabilité p .

Simulation (échantillon)

Pour obtenir un échantillon, il faudrait faire une expérience, c'est-à-dire vraiment lancer un grand nombre de fois 3 dés et noter les résultats qui apparaissent.

Les systèmes informatiques sont capables de produire des "*nombres pseudo-aléatoires*" qui ont les mêmes propriétés que s'ils avaient été "tirés au sort".

```
de3 = BinomialDistribution[3,  $\frac{1}{6}$ ];
```

Construisons un échantillon de 100 expériences

```
n = 100;
```

```
x = RandomInteger[de3, n]
```

```
{1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 2, 0, 0, 1, 1, 0, 0,
 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 2, 1, 1, 1, 0, 0,
 0, 0, 0, 0, 2, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 3, 0, 2, 0, 0, 2, 1,
 1, 1, 0, 0, 1, 2, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 2, 1, 0, 1, 0, 0, 1}
```

Plus généralement, un échantillon de taille n est une liste de n éléments noté

$$x = \{x_1, x_2, \dots, x_n\}$$

Paramètres empiriques

La moyenne empirique $m = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

```
m = Mean[x]; N[m]
```

```
0.54
```

L'écart-type empirique (non corrigé) $s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - m)^2}$

```
s = Sqrt[Apply[Plus, (x - m) ^ 2] / n]; N[s]
```

```
0.669627
```

Remarque: *Mathematica* utilise préférentiellement l'écart-type empirique corrigé

qui est un estimateur: $\hat{s} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2}$

```
N[StandardDeviation[x]]
```

```
0.673
```

Pour de grandes valeurs de n , la différence est infime, mais elle devient sensible pour de petites valeurs de n .

Référence: dernière page de https://www.deleze.name/marcel/sec2/applmaths/csud/statistique_2/2-stat_II.pdf

La fonction suivante, définie dans le package local *Statistique*, calcule l'écart-type empirique (non corrigé):

? StandardDeviationMLE

StandardDeviationMLE[data_] calcule l'écart-type de données brutes (échantillon).

```
s = StandardDeviationMLE[x]; N[s]
      |
      | valeur
```

0.669627

Groupement des données

Déterminons la liste des modalités

```
c = Range[0, 3]
      |
      | plage
```

{0, 1, 2, 3}

que nous noterons ici

$$C = \{c_1, c_2, c_3, c_4\}$$

Pour chaque valeur, déterminons l'effectif correspondant

```
effectifs = Map[Count[x, #] &, c]
      |
      | app· compte
```

{55, 37, 7, 1}

Dans cet exemple, les effectifs sont notés

$$\{n_1, n_2, n_3, n_4\}$$

Voici le tableau des effectifs

```
afficheTableau[{"Modalités : nombre de six en 3 dés",
  "Effectifs : nombre d'occurrences"}, None, {c, effectifs}]
      |
      | aucun
```

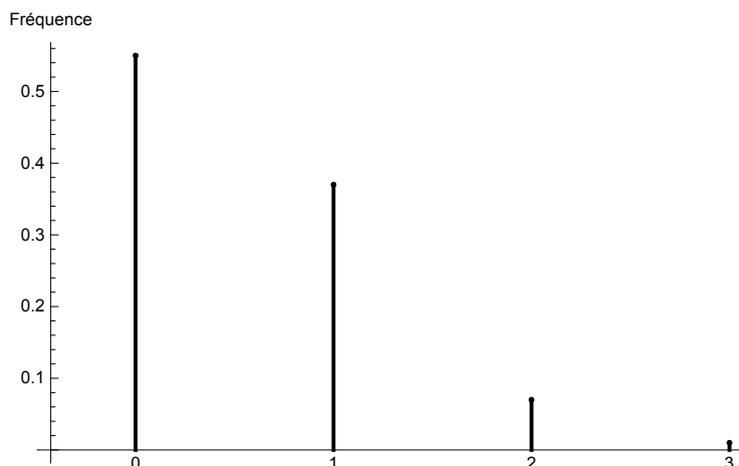
<i>Modalités : nombre de six en 3 dés</i>	0	1	2	3
<i>Effectifs : nombre d'occurrences</i>	55	37	7	1

Les fréquences

```
freq =  $\frac{\text{effectifs}}{n}$ 
```

$\left\{ \frac{11}{20}, \frac{37}{100}, \frac{7}{100}, \frac{1}{100} \right\}$

```
diagrammeBatons[c, freq, AxesLabel → {None, "Fréquence"}]
```

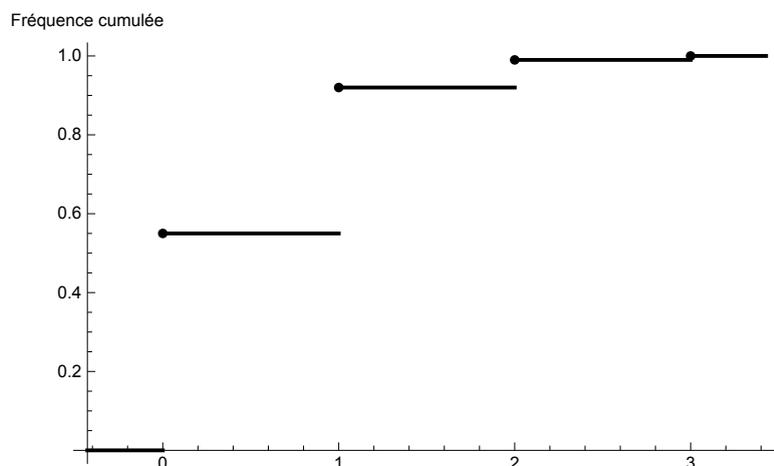


Les fréquences cumulées

```
freqCumulees = Accumulate[freq]
```

$$\left\{ \frac{11}{20}, \frac{23}{25}, \frac{99}{100}, 1 \right\}$$

```
diagrammeCumulatif[c, freq, AxesLabel → {None, "Fréquence cumulée"}]
```



Modèle statistique (distribution binomiale)

Dans l'exemple "Nombre de personnes par ménage" du § 1.1, la distribution théorique est inconnue.

Dans l'exemple courant "Nombre de six en 3 dés", la distribution théorique est connue. Il s'agit de la distribution binomiale $(3, \frac{1}{6})$.

Variable aléatoire

La variable étudiée ici est $X = \text{nombre de six obtenus en lançant 3 fois le dé}$.

X ne peut prendre que les valeurs $\{0, 1, 2, 3\} = \{c_0, c_1, c_2, c_3\}$

Probabilités

Le calcul des probabilités sera étudié en 4-ème année. Nous nous contentons de donner ici

quelques exemples introductifs.

La probabilité d'un événement représente sa "fréquence théorique".

Énumérons d'abord le *nombre de cas possibles*

$$\{1,1,1\}, \{1,1,2\}, \{1,1,3\}, \dots, \{6,6,5\}, \{6,6,6\} \quad \text{soit} \quad 6 \cdot 6 \cdot 6 = 216 \text{ cas.}$$

Énumérons maintenant le *nombre de cas comportant exactement 0 six*:

$$\{1,1,1\}, \{1,1,2\}, \{1,1,3\}, \dots, \{5,5,4\}, \{5,5,5\} \quad \text{soit} \quad 5 \cdot 5 \cdot 5 = 125 \text{ cas.}$$

Tous les cas étant équiprobables, la probabilité d'obtenir exactement 0 six est

$$p_0 = P \{X = 0\} = \frac{125}{216}$$

Énumérons maintenant le *nombre de cas comportant exactement 1 six*:

$$\{6,1,1\}, \{6,1,2\}, \{6,1,3\}, \dots, \{6,5,4\}, \{6,5,5\} \quad \text{soit} \quad 5 \cdot 5 = 25 \text{ cas,}$$

$$\{1,6,1\}, \{1,6,2\}, \{1,6,3\}, \dots, \{5,6,4\}, \{5,6,5\} \quad \text{soit} \quad 5 \cdot 5 = 25 \text{ cas,}$$

$$\{1,1,6\}, \{1,2,6\}, \{1,3,6\}, \dots, \{5,4,6\}, \{5,5,6\} \quad \text{soit} \quad 5 \cdot 5 = 25 \text{ cas,}$$

soit au total 75 cas.

Tous les cas étant équiprobables, la probabilité d'obtenir exactement 1 six est

$$p_1 = P \{X = 1\} = \frac{75}{216}$$

Énumérons maintenant le *nombre de cas comportant exactement 2 six*:

$$\{6,6,1\}, \{6,6,2\}, \{6,6,3\}, \{6,6,4\}, \{6,6,5\} \quad \text{soit} \quad 5 \text{ cas,}$$

$$\{6,1,6\}, \{6,2,6\}, \{6,3,6\}, \{6,4,6\}, \{6,5,6\} \quad \text{soit} \quad 5 \text{ cas,}$$

$$\{1,6,6\}, \{2,6,6\}, \{3,6,6\}, \{4,6,6\}, \{5,6,6\} \quad \text{soit} \quad 5 \text{ cas,}$$

soit au total 15 cas.

Tous les cas étant équiprobables, la probabilité d'obtenir exactement 2 six est

$$p_2 = P \{X = 2\} = \frac{15}{216}$$

Énumérons maintenant le *nombre de cas comportant exactement 3 six*:

$$\{6,6,6\} \quad \text{soit} \quad 1 \text{ cas}$$

Tous les cas étant équiprobables, la probabilité d'obtenir exactement 3 six est

$$p_3 = P \{X = 3\} = \frac{1}{216}$$

Mathematica connaît cette distribution.

`PDF[distr, cj]` signifie **P**robability **D**ensity **F**unction of **d**istr at **c**_j

`prob = Table[PDF[de3, cj], {cj, 0, 3}]`

`table` `[fonction de densité de probabilité]`

$$\left\{ \frac{125}{216}, \frac{25}{72}, \frac{5}{72}, \frac{1}{216} \right\}$$

Les probabilités ont les mêmes propriétés que les fréquences

$$0 \leq p_j \leq 1$$

$$p_0 + p_1 + p_2 + p_3 = 1$$

Plus généralement

$$p_j = P \{X = c_j\} = \text{probabilité de l'événement } X = c_j$$

$$0 \leq p_j \leq 1 \quad \text{et} \quad \sum_{j=0}^k p_j = 1$$

Espérance mathématique

L'espérance mathématique d'une variable aléatoire représente sa moyenne théorique. Rappelons que la moyenne empirique de données groupées se calcule comme suit

$$m = \bar{x} = \sum_{j=0}^k c_j f_j$$

L'espérance mathématique se calcule en remplaçant les fréquences par les probabilités correspondantes.

$$\mu = E(X) = \sum_{j=0}^k c_j p_j = c_0 p_0 + c_1 p_1 + c_2 p_2 + c_3 p_3 = 0 \frac{125}{216} + 1 \frac{75}{216} + 2 \frac{15}{216} + 3 \frac{1}{216} = \frac{1}{2} = 0.5$$

$\mu = \mathbf{c \cdot prob}$; $\mathbf{N}[\mu]$
|valeur

0.5

Variance et écart-type théoriques

Rappelons que la variance empirique de données groupées se calcule comme suit

$$v = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 = \frac{1}{n} \sum_{j=0}^k (c_j - m)^2 n_j = \sum_{j=0}^k (c_j - m)^2 \frac{n_j}{n} = \sum_{j=0}^k (c_j - m)^2 f_j$$

La variance théorique d'une variable aléatoire se calcule en remplaçant les fréquences par les probabilités correspondantes et la moyenne empirique par l'espérance mathématique:

$$V(X) = \sum_{j=0}^k (c_j - E(X))^2 p_j = \left(0 - \frac{1}{2}\right)^2 \frac{125}{216} + \left(1 - \frac{1}{2}\right)^2 \frac{75}{216} + \left(2 - \frac{1}{2}\right)^2 \frac{15}{216} + \left(3 - \frac{1}{2}\right)^2 \frac{1}{216} = \frac{5}{12}$$

L'écart-type théorique de la variable aléatoire X est

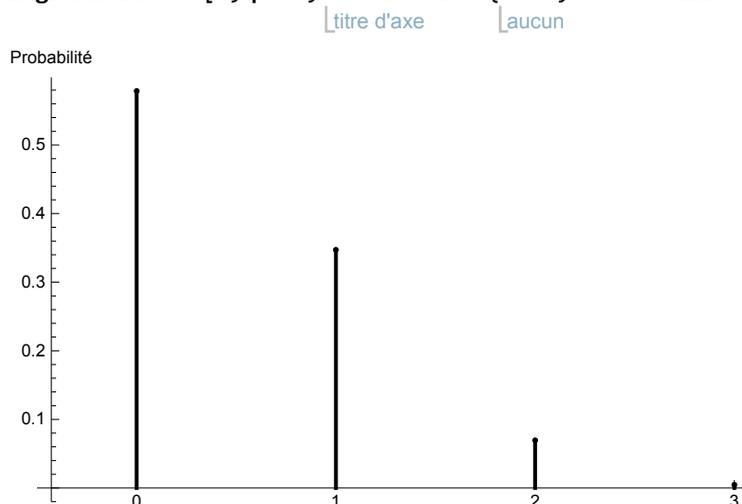
$$\sigma_X = \sqrt{V(X)} = \sqrt{\sum_{j=0}^k (c_j - E(X))^2 p_j} = \sqrt{\frac{5}{12}} \approx 0.645497$$

$\sigma = \sqrt{(c - \mu)^2 \cdot \mathbf{prob}}$; $\mathbf{N}[\sigma]$
|valeur

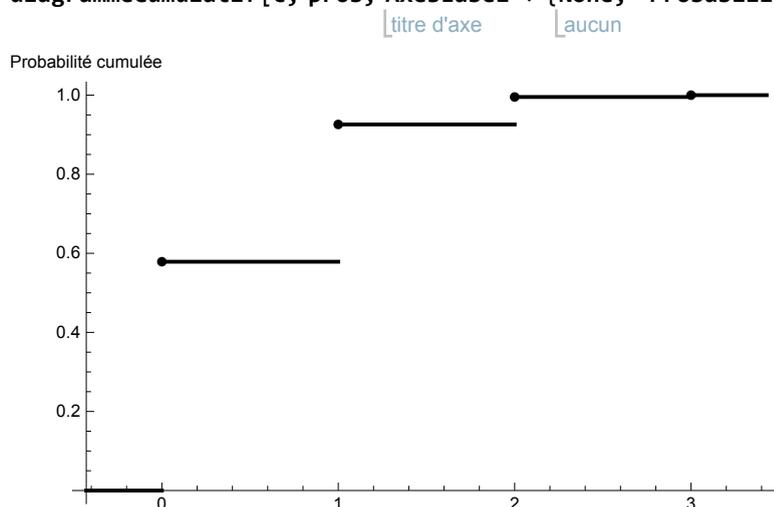
0.645497

Distribution

```
diagrammeBatons[c, prob, AxesLabel → {None, "Probabilité"}]
```



```
diagrammeCumulatif[c, prob, AxesLabel → {None, "Probabilité cumulée"}]
```



On remarquera que les diagrammes théoriques ont une forme semblable aux diagrammes empiriques.

Espérance et écart-type théorique de la distribution binomiale (k, p)

La formule qui exprime la probabilité de chaque issue sera étudié en quatrième année, au cours de mathématiques, dans le chapitre "*Calcul des probabilités*". En attendant, *Mathematica* nous en fournira les valeurs numériques comme il a été vu dans l'exemple précédent

```
Clear[k, p];
```

```
[efface
```

```
binomDistr = BinomialDistribution[k, p];
```

```
[distribution binomiale
```

$$p_i = \text{PDF}[\text{binomDistr}, i], \quad i = 0, 1, \dots, k.$$

Par contre, il nous sera utile de connaître dès maintenant l'espérance mathématique et l'écart-type théorique de la distribution binomiale (k, p) :

$\mu = \text{Mean}[\text{binomDistr}]$
└valeur moyenne

$k p$

$\sigma = \text{StandardDeviation}[\text{binomDistr}]$
└écart-type

$$\sqrt{k (1 - p) p}$$

Retenons le résultat : la probabilité de réussite de chaque expérience étant p , la moyenne et l'écart-type du nombre de réussites en k expériences indépendantes est

$$\mu = k p$$

$$\sigma = \sqrt{k p (1 - p)}$$

Dans le but de rendre plausibles ces deux formules, effectuons une simulation et comparons les valeurs empiriques et théoriques

$p = \frac{1}{6}$; $k = 50$; $n = 100\,000$;

$x = \text{RandomInteger}[\text{binomDistr}, n]$;
└entier aléatoire

$m = \text{Mean}[x]$; $N[m]$
└valeur m... └valeur

8.33635

$\mu = k p$; $N[\mu]$
└valeur

8.33333

$s = \text{StandardDeviationMLE}[x]$; $N[s]$
└valeur

2.64093

$\sigma = \sqrt{k p (1 - p)}$; $N[\sigma]$
└valeur

2.63523

Exercice 3.2 - 1 [Sans ordinateur]

Considérons le jeu suivant:

*On mise 7 Fr,
on lance 3 dés,
on gagne 6 Fr pour chaque six obtenu.*

- Calculez la probabilité des événements "obtenir exactement 0 six", "obtenir exactement 1 six", "obtenir exactement 2 six", "obtenir exactement 3 six".
- Pour la variable aléatoire $X =$ nombre de six obtenus, calculez l'espérance mathématique de X et l'écart-type théorique de X .
- Pour la variable aléatoire $Z =$ gain net d'une partie, dessinez le diagramme à bâtons des

probabilités (à la main).

- d) Pour la variable aléatoire $Z = \text{gain net d'une partie}$, calculez l'espérance mathématique de Z et l'écart-type théorique de Z .

Exercice 3.2 - 2 [Avec Mathematica]

Pour l'exemple du cours "Nombre de six en 3 dés", vérifiez empiriquement la loi des grands nombres de Bernoulli. Plus précisément,

- * formez un échantillon de taille n ;
- * comptez la fréquence de chaque modalité;
- * comparez les fréquences aux probabilités correspondantes;
- * répétez l'expérience pour une suite croissante de valeurs de n ;
- * concluez en reformulant la loi des grands nombres de Bernoulli.

Exercice 3.2 - 3 [Avec Mathematica]

On a lancé 500 fois une pièce de monnaie parfaitement équilibrée. Quelle est la probabilité que le nombre de faces soit situé entre 240 et 260 inclusivement ?

§ 3.3 Autres distributions

Distribution théorique donnée par des probabilités

Exercice 3.3 - 1

Dans chacune des situations suivantes, calculez les paramètres théoriques $E(X)$, σ_X [sans ordinateur] et comparez-les aux paramètres empiriques m , s obtenus au moyen d'une simulation [avec Mathematica]:

- a) $X = \text{nombre de face en lançant une fois une pièce de monnaie}$;
- b) $X = \text{nombre de six en lançant une fois un dé}$;
- c) X tel que
 - $X = 0$ avec prob. 0.6;
 - $X = 1$ avec prob. 0.3;
 - $X = 2$ avec prob. 0.1.

Indication pour c) : l'idée consiste à créer un échantillon uniformément distribué sur l'intervalle $[0; 1[$ auquel on applique la fonction suivante:

$$\begin{aligned} f(t) &= 0 \text{ pour } t < 0.6; \\ f(t) &= 1 \text{ pour } 0.6 \leq t < 0.9; \\ f(t) &= 2 \text{ pour } 0.9 \leq t. \end{aligned}$$

En *Mathematica*, cela donne

```
unif = UniformDistribution[{0, 1}];
      |distribution uniforme

n = 40;

Clear[f];
      |efface
```

```
f[t_] := Which[
  [quel]
  t < 0.6, 0,
  t < 0.9, 1,
  True, 2]
  [vrai]

Map[f, RandomReal[unif, n]]
[applique [nombre réel aléatoire]
{1, 1, 0, 2, 1, 0, 0, 0, 1, 0, 2, 0, 0, 0, 0, 2, 0,
0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 2, 0, 0, 0, 1}
```

Distribution géométrique

? GeometricDistribution

GeometricDistribution[p] represents a geometric distribution with probability parameter p . >>

Par exemple, une expérience aléatoire consistant à lancer un dé jusqu'à ce que l'on obtienne un six, le nombre de lancers avant d'obtenir le premier six est

$n = 40$;

```
jets = GeometricDistribution[ $\frac{1}{6}$ ];
  [distribution géométrique]
```

```
RandomInteger[jets, n]
```

[entier aléatoire]

```
{9, 4, 0, 13, 10, 2, 3, 6, 2, 18, 9, 5, 0, 0, 2, 5, 0, 1,
0, 0, 1, 2, 25, 4, 6, 0, 5, 16, 0, 5, 7, 5, 10, 0, 2, 6, 2, 1, 22, 2}
```

Si on désire que le dernier jet, qui est un six, soit compté, un échantillon est

```
x = 1 + RandomInteger[jets, n]
```

[entier aléatoire]

```
{1, 6, 8, 7, 11, 14, 10, 1, 13, 19, 1, 7, 21, 9, 7, 3, 4,
3, 8, 5, 5, 5, 2, 3, 8, 2, 3, 3, 3, 2, 3, 2, 5, 1, 7, 14, 2, 2, 5, 7}
```

Exercice 3.3 - 2 [Avec Mathematica]

On lance plusieurs fois un dé et on considère la variable aléatoire

$X = \text{nombre de lancers pour obtenir l'événement "six"}$.

Faites une simulation.

Dessinez le diagramme à bâtons empirique

Calculez la moyenne et l'écart-type empiriques.

Répétez plusieurs fois la simulation et observez les fluctuations des résultats.

Comparez les valeurs obtenues aux paramètres théoriques

$$E(X) = 6, \quad V(X) = 30$$

Distribution de Poisson

Nous voulons montrer que, lorsque k est grand et p petit, la distribution binomiale peut être approx-

imée par la distribution de Poisson de paramètre $\lambda = k p$

$k = 1000$; $p = 0.0001$; $\lambda = k p$

0.1

`binom = BinomialDistribution[k, p];`

[distribution binomiale]

`poisson = PoissonDistribution[λ];`

[distribution Poisson]

Calculons les probabilités des deux distributions pour les 10 premières modalités

`tabelle = Table[{c, PDF[binom, c], PDF[poisson, c]}, {c, 0, 9}]`

[table [fonction de densité... [fonction de densité de probabilité]

```
{ {0, 0.904833, 0.904837}, {1, 0.0904923, 0.0904837},
  {2, 0.00452054, 0.00452419}, {3, 0.000150398, 0.000150806},
  {4, 3.74906 × 10-6, 3.77016 × 10-6}, {5, 7.46887 × 10-8, 7.54031 × 10-8},
  {6, 1.23871 × 10-9, 1.25672 × 10-9}, {7, 1.75915 × 10-11, 1.79531 × 10-11},
  {8, 2.18376 × 10-13, 2.24414 × 10-13}, {9, 2.40723 × 10-15, 2.49349 × 10-15}}
```

`afficheTableau[None, {"Modalité", "Binomial", "Poisson"}, tabelle]`

[aucun]

[binôme]

Modalité	Binomial	Poisson
0	0.904833	0.904837
1	0.0904923	0.0904837
2	0.00452054	0.00452419
3	0.000150398	0.000150806
4	3.74906×10^{-6}	3.77016×10^{-6}
5	7.46887×10^{-8}	7.54031×10^{-8}
6	1.23871×10^{-9}	1.25672×10^{-9}
7	1.75915×10^{-11}	1.79531×10^{-11}
8	2.18376×10^{-13}	2.24414×10^{-13}
9	2.40723×10^{-15}	2.49349×10^{-15}

L'expression mathématique de la distribution binomiale - qui dépend de deux paramètres (k , p) - ne sera étudiée que plus tard, au cours de mathématiques en 4^{ème} année. Par contre, la forme mathématique de la distribution de Poisson - qui ne dépend que d'un seul paramètre λ - est plus simple : les probabilités des modalités 0, 1, 2, 3, 4 sont

$$p_0 = e^{-\lambda}$$

$$p_1 = \lambda p_0 = \lambda e^{-\lambda}$$

$$p_2 = \frac{\lambda}{2} p_1 = \frac{\lambda^2}{2} e^{-\lambda}$$

$$p_3 = \frac{\lambda}{3} p_2 = \frac{\lambda^3}{6} e^{-\lambda}$$

$$p_4 = \frac{\lambda}{4} p_3 = \frac{\lambda^4}{24} e^{-\lambda}$$

Plus généralement, pour la distribution de Poisson de paramètre λ , la probabilité de la modalité j est

$$p_j = P(X = j) = \frac{\lambda^j}{j!} e^{-\lambda}, \quad j \in \mathbb{N}$$

Pour que l'approximation soit satisfaisante, il faut exiger que k soit assez grand et

p soit suffisamment petit;
plus précisément,

$$k \geq 50 \quad \text{et} \quad p < \frac{5}{k}.$$

Exercice 3.3 - 3 [Avec Mathematica]

La probabilité qu'un individu ait une mauvaise réaction à l'injection d'un certain sérum est de 0.001. Déterminez la probabilité pour que, sur 2000 individus, plus de deux individus aient une réaction dangereuse.

Directives et indications :

- * considérez la distribution de Poisson qui approxime la distribution binomiale;
- * dessinez le diagramme à bâtons des probabilités;
- * considérez l'événement complémentaire "0, 1 ou 2 individu(s) réagi(ssen)t dangereusement";
- * calculez la probabilité de l'événement à partir de la fonction de distribution.

§ 3.4 Propriétés de l'espérance mathématique et de la variance

Exercice 3.4 - 1 [Sans ordinateur]

Considérons l'échantillon suivant dont les éléments représentent des nombres de bouteilles achetées par des clients

$$x = \{24, 6, 12, 2\};$$

On s'intéresse maintenant aux prix correspondants selon le tarif suivant: 7 francs par bouteille plus 3 francs pour les frais de port et d'emballage. On obtient ainsi un nouvel échantillon

$$y = 7x + 3$$

$$\{171, 45, 87, 17\}$$

Comparez les moyennes et variances des deux échantillons :

$$\begin{array}{ccc} m(y) & \text{et} & 7m(x) + 3 \\ v(y) & \text{et} & 7v(x) + 3 \end{array}$$

Comment $v(y)$ dépend-t-il des frais de port et d'emballage ?

A quelle grandeur se rapportant à l'échantillon x faut-il comparer $v(y)$?

Premières propriétés de la moyenne et de l'écart-type empirique

Généralisons: si x désigne un échantillon et a, b des constantes réelles, la moyenne arithmétique m et la variance v vérifient

$$\begin{array}{l} m(a * x + b) = a * m(x) + b \\ v(a * x + b) = a^2 * v(x) \end{array}$$

Démonstration

$$m(y) = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (a x_i + b) = \frac{a}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n b = a * m(x) + b$$

$$v(y) = \frac{1}{n} \sum_{i=1}^n (y_i - m(y))^2 = \frac{1}{n} \sum_{i=1}^n (a x_i + b - (a * m(x) + b))^2 =$$

$$\frac{1}{n} \sum_{i=1}^n (a x_i - a * m(x))^2 = a^2 \frac{1}{n} \sum_{i=1}^n (x_i - m(x))^2 = a^2 v(x)$$

Premières propriétés de l'espérance et de l'écart-type théorique

A chaque paramètre empirique (= de l'échantillon) correspond un paramètre théorique (= de la population). Le tableau suivant donne les principales correspondances

Notion empirique	Symbole empirique	Symbole théorique	Notion théorique
échantillon	x	X	variable aléatoire
fréquence	f _i	p _i	probabilité
fréquence cumulée	F	Φ	distribution
densité empirique	h	φ	densité théorique
moyenne	m=m(x)	μ=E(X)	espérance mathématique
variance empirique	s ² =v(x)	σ ² =V(X)	variance théorique

La variable aléatoire sera notée X. Dans l'exemple du § 1, la variable X représente le nombre de personnes par ménage pour toute la population.

Considérons par exemple le jeu suivant:

On mise 7 Fr,
on lance 3 dés,
on gagne 6 Fr pour chaque six obtenu.

En fonction du nombre de six obtenu, on peut dresser la table des gains nets

pour 0 six, le gain net est de $0*6 - 7 = -7$
pour 1 six, le gain net est de $1*6 - 7 = -1$
pour 2 six, le gain net est de $2*6 - 7 = 5$
pour 3 six, le gain net est de $3*6 - 7 = 11$

Une partie consistant à lancer 3 fois le dé, si on appelle

X = nombre de six et

Y = gain net,

on a la relation

$$Y = 6X - 7$$

Nous avons vu dans le § 3.1 que

$$E(X) = \sum_{j=0}^k x_j p_j = 0 \frac{125}{216} + 1 \frac{75}{216} + 2 \frac{15}{216} + 3 \frac{1}{216} = \frac{1}{2} \quad (\text{les } x_j \text{ désignent ici les modalités de } X)$$

$$V(X) =$$

$$\sum_{j=0}^k (x_j - E(X))^2 p_j = \left(0 - \frac{1}{2}\right)^2 \frac{125}{216} + \left(1 - \frac{1}{2}\right)^2 \frac{75}{216} + \left(2 - \frac{1}{2}\right)^2 \frac{15}{216} + \left(3 - \frac{1}{2}\right)^2 \frac{1}{216} = \frac{5}{12}$$

Calculons de même

$$E(Y) =$$

$$\sum_{j=0}^k y_j p_j = (-7) \frac{125}{216} + (-1) \frac{75}{216} + 5 \frac{15}{216} + 11 \frac{1}{216} = -4 \quad (\text{les } y_j \text{ désignent ici les modalités de } Y)$$

$$V(Y) =$$

$$\sum_{j=0}^k (y_j - E(Y))^2 p_j = (-7+4)^2 \frac{125}{216} + (-1+4)^2 \frac{75}{216} + (5+4)^2 \frac{15}{216} + (11+4)^2 \frac{1}{216} = 15$$

Par ailleurs, les expressions suivantes donnent les mêmes résultats

$$6E(X) - 7 = 6 \times \frac{1}{2} - 7 = -4$$

$$6^2 V(X) = 6^2 \frac{5}{12} = 15$$

et on a la proposition

$E(a * X + b) = a * E(X) + b$ $V(a * X + b) = a^2 * V(X)$

Démonstration

Ici, les valeurs des variables aléatoires X , Y sont notées x_j , y_j respectivement.

$$E(Y) = \sum_{j=0}^k y_j p_j = \sum_{j=0}^k (a x_j + b) p_j = a \sum_{j=0}^k x_j p_j + b \sum_{j=0}^k p_j = a * E(X) + b$$

$$V(Y) = \sum_{j=0}^k (y_j - E(Y))^2 p_j = \sum_{j=0}^k (a x_j + b - (a * E(X) + b))^2 p_j =$$

$$\sum_{j=0}^k (a x_j - a * E(X))^2 p_j = a^2 \sum_{j=0}^k (x_j - E(X))^2 p_j = a^2 V(X)$$

Exercice 3.4 - 2 [Avec Mathematica]

Données de l'exercice:

https://www.deleze.name/marcel/sec2/applmaths/csud/statistique_1/annexes/3-4-2-donnees-exercice.nb

Directives : pour cet exercice, on pourra utiliser les fonctions `Mean[.]`, `StandardDeviationMLE[...]`, etc.

La puissance électrique d'une ampoule étant de 60 W, on considère maintenant l'échantillon des consommations en kWh, c'est-à-dire $y = 0.06 x$

Calculez $m(y)$, $s(y)$.

Comparez $m(0.06 x)$ et $0.06 m(x)$
 $s(0.06 x)$ et $0.06 s(x)$

Exercice 3.4 - 3 [Sans ordinateur]

Dans le § 3.3, on a utilisé la distribution géométrique de paramètre $\frac{1}{6}$ dont on connaît l'espérance et la variance théorique

```
jets = GeometricDistribution[ $\frac{1}{6}$ ];
      [distribution géométrique]
```

```
Mean[jets]
      [valeur moyenne]
```

5

```
Variance[jets]
      [variance]
```

30

On demande d'en déduire l'espérance et l'écart-type théorique de la variable aléatoire
 $X =$ nombre de lancers du dé jusqu'à l'obtention du premier six.

Exercice 3.4 - 4 [Sans ordinateur]

Considérons un échantillon dont les éléments sont des nombres de bouteilles achetées par des clients

$$x = \{24, 6, 12, 2\};$$

Formez les échantillons suivants, appelés respectivement

les carrés;

les écarts à la moyenne ou l'échantillon de la variable centrée;

les carrés des écarts à la moyenne ou l'échantillon du carré de la variable centrée:

$$\begin{aligned} & x^2 \\ & x - m(x) \\ & (x - m(x))^2 \end{aligned}$$

Calculez et comparez

$$\begin{aligned} & s^2(x) \\ & m((x - m(x))^2) \\ & m(x^2) - m^2(x) \end{aligned}$$

Autres propriétés de la variance empirique

Généralisons. On a la relation

$$s^2(x) = m((x - m(x))^2)$$

En mots :

La variance empirique est la moyenne des carrés des écarts à la moyenne

Démonstration. Posons

$$z_i = (x_i - m(x))^2 \quad \text{et} \quad z = (x - m(x))^2$$

On a alors

$$s^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - m(x))^2 = \frac{1}{n} \sum_{i=1}^n z_i = m(z) = m((x - m(x))^2)$$

Généralisons. On a la relation

$$s^2(x) = m(x^2) - m^2(x)$$

En mots :

La variance empirique est égale à la moyenne des carrés moins le carré de la moyenne

Démonstration. Posons

$$z_i = (x_i - m(x))^2 \quad \text{et} \quad z = (x - m(x))^2 = x^2 - 2m(x)x + m^2(x)$$

En utilisant les règles de calcul des paragraphes 3.1 à 3.3, on a

$$\begin{aligned} s^2(x) &= m(z) = m(x^2 - 2m(x)x + m^2(x)) \\ &= m(x^2) - 2m(x)m(x) + m^2(x) = m(x^2) - m^2(x) \end{aligned}$$

Autres propriétés de la variance théorique

Considérons par exemple le jeu suivant:

on lance 2 dés,

on observe le nombre de six obtenus.

Calculons d'abord l'espérance mathématique et la moyenne

$$E(X) = 0 \frac{25}{36} + 1 \frac{10}{36} + 2 \frac{1}{36} = \frac{1}{3}$$

$$V(X) = \left(0 - \frac{1}{3}\right)^2 \frac{25}{36} + \left(1 - \frac{1}{3}\right)^2 \frac{10}{36} + \left(2 - \frac{1}{3}\right)^2 \frac{1}{36} = \frac{5}{18}$$

D'une part, la variance est égale à la moyenne des carrés des écarts. Dans une terminologie mieux adaptée aux paramètres théoriques, on dit

$X - E(X)$ est la variable centrée

$(X - E(X))^2$ est le carré de la variable centrée

$E((X - E(X))^2)$ est l'espérance mathématique du carré de la variable centrée.

$$Z = (X - E(X))^2$$

$$z_1 = \left(0 - \frac{1}{3}\right)^2, z_2 = \left(1 - \frac{1}{3}\right)^2, z_3 = \left(2 - \frac{1}{3}\right)^2$$

$$E(Z) = z_1 \frac{25}{36} + z_2 \frac{10}{36} + z_3 \frac{1}{36} = D^2(X)$$

D'autre part, la variance est égale à l'espérance du carré moins le carré de l'espérance

$$E(X^2) = 0^2 \frac{25}{36} + 1^2 \frac{10}{36} + 2^2 \frac{1}{36} = \frac{7}{18}$$

$$E(X^2) - E^2(X) = \frac{7}{18} - \left(\frac{1}{3}\right)^2 = \frac{5}{18} = V(X)$$

Plus généralement

$$\begin{aligned} V(X) &= E((X - E(X))^2) \\ V(X) &= E(X^2) - E^2(X) \end{aligned}$$

En mots :

La variance théorique est l'espérance mathématique du carré de la variable centrée.

La variance théorique est égale à l'espérance du carré moins le carré de l'espérance.

Démonstration

La première relation peut être considérée comme une définition de la variance.

La deuxième relation se déduit des règles de calcul des paragraphes 3.1 à 3.3

$$\begin{aligned} V(X) &= E((X - E(X))^2) = E(X^2 - 2XE(X) + E^2(X)) \\ &= E(X^2) - 2E(X)E(X) + E^2(X) = E(X^2) - E^2(X) \end{aligned}$$

Exercice 3.4 - 5 [Avec Mathematica]

Reprenons les données numériques de l'exercice 3.4-2:

https://www.deleze.name/marcel/sec2/applmaths/csud/statistique_1/annexes/3-4-2-donnees-exercice.nb

Directives : pour cet exercice, on pourra utiliser les fonctions `Mean[.]`, `StandardDeviationMLE[...]`, etc.

Calculez $m(x^2)$ et $m^2(x)$
 et comparez $s^2(x)$ et $m(x^2) - m^2(x)$.

Exercice 3.4 - 6 [Sans ordinateur]

- a) On mise 5 Fr. On lance une pièce de monnaie.
 Sur pile, on gagne 8 Fr; sur face 1 Fr (gains bruts).
 Notons A la variable aléatoire du gain net d'une partie.
 Calculez $E(A)$ et $V(A)$.
- b) On mise 8 Fr. On lance un dé.
 On gagne 2 Fr par point indiqué par le dé (gain brut).
 Notons B la variable aléatoire du gain net d'une partie.
 Dessinez le diagramme à bâtons des probabilités de la variable aléatoire B .
 Calculez $E(B)$, $V(B)$, $E(B^2)$, $E((B - E(B))^2)$
 Comparez $V(B)$, $E((B - E(B))^2)$, $E(B^2) - E^2(B)$

Liens

Vers les corrigés des exercices :

https://www.deleze.name/marcel/sec2/applmaths/csud/corriges/statistique_1/3-stat_I-cor.pdf

Vers la page mère Applications des mathématiques

<https://www.deleze.name/marcel/sec2/applmaths/csud/index.html>