Applications des mathématiques

Statistique II

Modèles à plusieurs variables
Variables indépendantes

Si X et Y sont des variables aléatoires **indépendantes**,
d'écarts-types respectifs σ_X, σ_Y,
alors l'écart-type de la somme est

$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$$

Version pour *Mathematica*
Edition 2017
Marcel Délèze
https://www.deleze.name/marcel/sec2/applmaths/csud/index.html

§ 1 Modèles à plusieurs variables

§ 1.1 Somme de deux variables statistiques

$m(x+y) = m(x) + m(y)$

$E(X+Y) = E(X) + E(Y)$

Exercice 1.1 - 1

Considérons par exemple des couples dont on mesure les masses respectives. Notons

x_i =masse de l'homme du couple numéro i,

y_i =masse de la femme du couple numéro i,

$c_i = (x_i, \; y_i)$ = masses du couple numéro i (il s'agit d'un échantillon bidimentionnel);

$z_i = x_i + \; y_i$ = somme des masses du couple numéro i.

Etant donné l'échantillon de dimension 2 suivant

```
c = {(75, 61), (68, 65), (85, 62), (72, 78)}
```

on peut en tirer l'échantillon marginal du poids des hommes (c'est-à-dire une *composante* de l'échantillon multidimentionnel)

```
x = {75, 68, 85, 72},
```

l'échantillon marginal du poids des femmes

```
y = {61, 65, 62, 78}
```

et la somme des masses des couples (remarquez que pour sommer deux échantillons, il est nécessaire qu'ils soient de même taille)

```
z = {136, 133, 147, 150}
```

Comparez

la moyenne des sommes $m(x+y) = m(z)$;

la somme des moyennes $m(x) + m(y)$.

Moyenne empirique d'une somme

Généralisons. La formule suivante est vraie : pour deux échantillons x, y de même taille n, on a

$$\boxed{m(x+y) = m(x) + m(y)}$$

En mots :

La moyenne d'une somme est égale à la somme des moyennes

Démonstration

$$m(x+y) = \frac{1}{n}\sum_{i=1}^{n}(x_i + y_i) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i\right) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} y_i = m(x) + m(y)$$

Espérance mathématique d'une somme

On dispose par exemple d'une paire de dés que l'on lance; le premier est rouge et le deuxième est vert. Les variables aléatoires correspondantes sont notées comme suit

X = nombre obtenu sur le dé rouge,

Y = nombre obtenu sur le dé vert,

$Z = X + Y$ = somme des points obtenus sur les deux dés.

Pour chacun des deux dés, on a

$$p(X=1) = \frac{1}{6}, \quad \ldots, \quad p(X=6) = \frac{1}{6}$$
$$p(Y=1) = \frac{1}{6}, \quad \ldots, \quad p(Y=6) = \frac{1}{6}$$

Calculons l'espérance mathématique du nombre obtenu en lançant un dé

$$E(X) = 1\,\frac{1}{6} + 2 \times \frac{1}{6} + 3\,\frac{1}{6} + 4\,\frac{1}{6} + 5\,\frac{1}{6} + 6\,\frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$
$$E(Y) = 1\,\frac{1}{6} + 2 \times \frac{1}{6} + 3\,\frac{1}{6} + 4\,\frac{1}{6} + 5\,\frac{1}{6} + 6\,\frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

Calculons maintenant les probabilités associées aux valeurs de la somme

$$p(Z=2) = p((X, Y) = (1, 1)) = \frac{1}{36}$$
$$p(Z=3) = p((X, Y) \in \{(1, 2), (2, 1)\}) = \frac{2}{36}$$
$$p(Z=4) = p((X, Y) \in \{(1, 3), (3, 1), (2, 2)\}) = \frac{3}{36}$$
$$p(Z=5) = p((X, Y) \in \{(1, 4), (4, 1), (2, 3), (3, 2)\}) = \frac{4}{36}$$
$$p(Z=6) = p((X, Y) \in \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\}) = \frac{5}{36}$$
$$p(Z=7) = p((X, Y) \in \{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\}) = \frac{6}{36}$$
$$p(Z=8) = p((X, Y) \in \{(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)\}) = \frac{5}{36}$$
$$p(Z=9) = p((X, Y) \in \{(3, 6), (6, 3), (4, 5), (5, 4)\}) = \frac{4}{36}$$
$$p(Z=10) = p((X, Y) \in \{(4, 6), (6, 4), (5, 5)\}) = \frac{3}{36}$$
$$p(Z=11) = p((X, Y) \in \{(5, 6), (6, 5)\}) = \frac{2}{36}$$
$$p(Z=12) = p((X, Y) = (6, 6)) = \frac{1}{36}$$

Puis calculons l'espérance mathématique de la somme avec les notations généralisées suivantes:

$$k = 11;\ z_1 = 2;\ \ldots;\ z_k = 12$$
$$p_1 = P(Z = z_1) = \frac{1}{36};\ \ldots;\ p_k = P(Z = z_k) = \frac{1}{36}$$
$$E(Z) = \sum_{j=1}^{k} z_j\, p_j = 2 \times \frac{1}{36} + 3\,\frac{2}{36} + 4\,\frac{3}{36} + 5\,\frac{4}{36} + 6\,\frac{5}{36} +$$
$$7\,\frac{6}{36} + 8\,\frac{5}{36} + 9\,\frac{4}{36} + 10\,\frac{3}{36} + 11\,\frac{2}{36} + 12\,\frac{1}{36} = 7$$

On a les relations

$$X + Y = Z$$

$$E(X) + E(Y) = \frac{7}{2} + \frac{7}{2} = 7 = E(Z)$$

Plus généralement,

$$\boxed{E(X+Y) = E(X) + E(Y)}$$

(sans démonstration)
En mots :

$$\boxed{\text{L'espérance d'une somme est égale à la somme des espérances}}$$

La relation est aussi valide pour des variables aléatoires X, Y qui ont des distributions différentes.
La relation est aussi valide pour des variables aléatoires X, Y continues.

Exercice 1.1 - 2

Soient $X_1, X_2, \ldots, X_n$ des variables de Bernoulli de même distribution $P\{X_j = 0\} = p$, $P\{X_j = 1\} = 1 - p = q$; notons $S = X_1 + X_2 + \ldots + X_n$ la variable binomiale $(n,\ p)$. Calculez l'espérance de S.

§ 1.2 Propriétés non générales

Nous nous intéressons à des propriétés qui - sous certaines hypothèses - sont vérifiées mais qui - en général - ne sont pas valides

$$m(x\,y) \overset{?}{=} m(x)\,m(y)$$
$$v(x+y) \overset{?}{=} v(x) + v(y)$$

En mots :

la moyenne d'un produit est-elle égale au produit des moyennes ?
la variance d'une somme est-elle égale à la somme des variances ?

Exercice 1.2 - 1

Considérons par exemple des couples dont on mesure les masses respectives. Notons

x_i =masse de l'homme du couple numéro *i*,
y_i =masse de la femme du couple numéro *i*,
$c_i = (x_i,\ y_i)$ = masses du couple numéro *i*.

Etant donné l'échantillon **à deux dimensions** suivant

```
c = {{75, 61}, {68, 65}, {85, 62}, {72, 78}};
```

on peut en tirer l'échantillon marginal du poids des hommes

```
x = {75, 68, 85, 72}
```

l'échantillon marginal du poids des femmes

```
y = {61, 65, 62, 78}
```

la somme des masses

```
x + y = {136, 133, 147, 150}
```

et le produit des masses (remarquez que pour multiplier deux échantillons, il est nécessaire qu'ils soient de même taille)

```
x y = {4575, 4420, 5270, 5616}
```

a) Comparez la moyenne des produits $m(x\,y)$ avec le produit des moyennes $m(x)\,m(y)$.

b) Comparez la variance de la somme $v(x+y)$ avec la somme des variances $v(x) + v(y)$.

Covariance

En général, la moyenne du produit n'est pas égale au produit des moyennes. La différence est appelée **covariance:**

$$\text{cov}\,(x, y) = m\,(x\,y) - m\,(x)\,m\,(y)$$

Variance de la somme et somme des variances

En général, la variance de la somme n'est pas égale à la somme des variances.

Espérance du produit et variance théorique d'une somme

Pour les valeurs théoriques, on a une situation semblable. En général, on doit distinguer, d'une part

l' espérance du produit	$E\,(X\,Y)$
le produit des espérances	$E\,(X)\,E\,(Y)$

d'autre part

la variance d' une somme	$V\,(X+Y)$
la somme des variances	$V\,(X) + V\,(Y)$

Nous illustrerons ces situations par des exemples dans les paragraphes 2.2, 2.3 et 2.4.

Exercice 1.2 - 2 [Avec *Mathematica*]

<u>Directives</u> : pour cet exercice, on pourra utiliser le package *Statistics`DescriptiveStatistics`* où sont définies les fonctions *Mean[..]*, *StandardDeviationMLE[...]*, etc.

Pour télécharger les données numériques:

https://www.deleze.name/marcel/sec2/applmaths/csud/statistique_2/annexes/1-2-2-donnees-exercice.nb

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

Exercice 1.2 - 3 [Sans ordinateur]

<u>Données</u> : on dispose de deux lots d'ampoules électriques, 10 ampoules provenant d'un producteur A et 10 ampoules provenant du producteur B. On mesure les durées de vie de chaque ampoule. Les durées de vie sont exprimées en heures et arrondies à l'entier.

```
a = {1370, 1347, 1677, 1832, 1297, 1008, 2018, 2117, 1260, 1019};
```

```
b = {792, 1039, 1583, 2004, 1184, 1683, 1975, 1690, 1712, 1684};
```

Les échantillons précédents sont à interpréter comme suit:

la première ampoule du producteur A a une durée de vie de 1370 h,
la deuxième ampoule du producteur A a une durée de vie de 1347 h,
....
la première ampoule du producteur B a une durée de vie de 792 h,
la deuxième ampoule du producteur B a une durée de vie de 1039 h,
...

Pour interpréter la somme des deux échantillons, on peut imaginer la procédure suivante:
on dispose de 10 supports de lampes;
au départ, chaque support est muni d'une ampoule du producteur A
et, en réserve, d'une ampoule du producteur B.
Pendant combien d'heures chaque support de lampe pourra-t-il éclairer ?

```
z = a + b
{2162, 2386, 3260, 3836, 2481, 2691, 3993, 3807, 2972, 2703}
```

Directives : pour cet exercice, il est demandé d'utiliser des formulas explicites pour calculer les moyennes et les écart-types. En particulier, on n'utilisera ni le package *Statistics`DescriptiveStatistics`*, ni les fonctions statistiques préprogrammées de votre calculatrice de poche.

a)	Calculez	les moyennes	$m(a)$,	$m(b)$,	$m(a+b)$
		les écart-types	$s(a)$,	$s(b)$,	$s(a+b)$
b)	Comparez	les moyennes	$m(a)+m(b)$	et	$m(a+b)$
		les écart-types	$s(a)+s(b)$	et	$s(a+b)$
		les variances	$v(a)+v(b)$	et	$v(a+b)$

Exercice 1.2 - 4 [Sans ordinateur]

Jeu 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.

Jeu 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.

Jeu C: On mise 13 Fr. On lance une pièce de monnaie et un dé.
Sur pile, on gagne 8 Fr; sur face 1 Fr; de plus,
on gagne 2 Fr par point indiqué par le dé (gains bruts).
Notons C la variable aléatoire du gain net d'une partie (il s'agit de la somme des jeux A et B).

Calculez
$E(C)$ et $V(C)$.
Comparez les espérances mathématiques $E(A+B)$ et $E(A)+E(B)$.
Comparez les variances $V(A+B)$ et $V(A)+V(B)$.

Lien vers les corrigés des exercices

https://www.deleze.name/marcel/sec2/applmaths/csud/corriges/statistique_2/1-stat_II-cor.pdf