# T.D. de Probabilités et Statistiques

## Licence Informatique

Gilles HUNAULT, 1998

## T.D. 3 : Variables aléatoires et Corrélation

1. Non additivité de la variance
2. Diverses v.a. (somme, produit...) pour 2 dés à 3 faces
3. Lois et valeurs comme $V < m$ pour $\mathcal{B}(n, p)$
4. Loi de *Bernoulli* généralisée
5. Centrage et Réduction
6. Qu'est-ce qu'une moyenne ?

# 1. Non additivité de la variance

### Enoncé

Peut-on inventer une v.a. $X$ avec 2 valeurs, une v.a. $Y$ avec 3 valeurs et montrer facilement que $V(X+Y) \neq V(X)+V(Y)$ ?

Peut-on facilement trouver un cas où $V(X+Y) = V(X)+V(Y)$ ?

### Solution

On ne peut pas prendre des valeurs $x_1$ et $x_2$ pour $X$ avec des probabilités respectives $p_{X,1}$ et $p_{X,2}$, des valeurs $y_1$, $y_2$ et $y_3$ $Y$ avec comme probabilité respective $p_{Y,1}$, $p_{Y,2}$ et $p_{Y,3}$ car ne sachant pas comment $X$ et $Y$ sont construites (c'est à dire sur quels évènements elles reposent), on ne peut pas définir $X+Y$.

Si $Y$ est constante, alors $V(Y) = 0$ et $V(X+Y) = V(X)+V(Y) = V(X)$.

# 2. Diverses v.a. pour 2 dés à 3 faces

### Enoncé

On lance 2 dés à 3 faces. Etudiez les variables $S$, $P$, $M = S * P$, $D$ qui correspondent respectivement à la somme, le produit, au produit de $S$ par $P$ et à la valeur absolue de la différence des chiffres inscrits sur le dé.

### Solution

Avant se lancer directement dans les calculs (la somme $S$ vaut 2 pour 1+1 seulement, elle vaut 3 pour 1+2 et 2+1, elle vaut 4 pour 1+3,2+2, 3+1...) il faut essayer de voir comment on peut structurer l'ensemble de départ. Avec $n$ faces pour le premier dé et $n$ faces pour le second, il y a $n^2$ couples de résultats possibles, même si les dés ne sont pas discernables. Ainsi $(1,2)$ et $(2,1)$ sont deux évènements élementaires, alors qu'on ne les identifier que si par exemple les dés sont de couleur différente. Le système complet d'évènements est alors l'ensemble des couples $(i,j)$ pour $i$ et $j$ de 1 à $n$, chaque couple équiprobable ayant une probabilité de $1/36$.

Une autre solution aurait consisté à ne retenir que les couples $(x,y)$ discernables, en les ordonnant par exemple par $x \leq y$. Mais on alors on aurait moins de couples (combien ?) et ils ne seraient plus équiprobables. Pour $n = 3$, au lieu des 9 couples

(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)

avec chacun comme probabilité 1/9, on n'aurait plus que les 6 couples

(1,1),(1,2),(1,3),(2,2),(2,3),(3,3)

de probabilité respective

1/9,2/9,2/9,1/9,2/9,1/9

Le plus simple pour construire tous les cas possibles est d'écrire un programme pour gérer tous les calculs. La partie remplissage du tableau des données, noté svd (comme série de valeurs), peut se faire par l'algorithme suivant où col_S désigne la colonne des valeurs de $S$, col_P celle de $P$ etc. Une implémentation simple consiste à prendre col_S=1, col_P=2 etc.

On trouvera en annexe un programme complet écrit en *tcl* qui effectue tous les calculs.

```
i <-- 1
v <-- 0
tant que i <= n
  j <-- 1
  tant que j <= n
      v <-- v + 1
      sdv[v,col_S] <-- (i + j)
      sdv[v,col_P] <-- (i * j)
      sdv[v,col_M] <-- (i * j)*(i + j)
      sdv[v,col_D] <-- abs(i - j)
      j <-- j + 1
  fin tant que sur j
  i <-- i + 1
fin tant que sur i
```

Ce programme permet de gérer aisément deux dés à $n$ faces. L'énoncé demandait d'étudier le cas $n = 3$, ce qui fournit les résultats suivants.

```
Etude de 2 ds avec chacun 3 faces
... On passe en revue les 9 cas    possibles.

tude de la variable : S = somme des valeurs

   S  |    1    2    3
 -----+--------------------------------
   1  |    2    3    4
   2  |    3    4    5
   3  |    4    5    6

   S  |    2    3    4    5    6
 -----+--------------------------
  9p  |    1    2    3    2    1    total : 9

variable S moyenne    :     4.000  (somme des valeurs :  36)
           variance   :     1.333  (somme des carrss : 156)
           cart-type :     1.155
           cdv        :    28.868 % (coefficient de variation)

tude de la variable : P = produits des valeurs

   P  |    1    2    3
 -----+--------------------------------
   1  |    1    2    3
   2  |    2    4    6
   3  |    3    6    9

   P  |    1    2    3    4    6    9
 -----+--------------------------------
  9p  |    1    2    2    1    2    1    total : 9

variable P moyenne    :     4.000 (somme des valeurs :   36)
           variance   :     5.778 (somme des carrs  :  196)
           cart-type :     2.404
           cdv        :    60.093 % (coefficient de variation)
```

```
tude de la variable : M = somme*produit des valeurs

   M  |    1    2    3
 -----+--------------------------------
   1  |    2    6   12
   2  |    6   16   30
   3  |   12   30   54

   M  |    2    6   12   16   30   54
 -----+--------------------------------
  9p  |    1    2    2    1    2    1    total : 9

variable M moyenne    :    18.667 (somme des valeurs :    168)
           variance   :   244.444 (somme des carrs  :   5336)
           cart-type :    15.635
           cdv        :    83.757 % (coefficient de variation)

tude de la variable : D = valeur absolue de la diffrence

   D  |    1    2    3
 -----+--------------------------------
   1  |    0    1    2
   2  |    1    0    1
   3  |    2    1    0

   D  |    0    1    2
 -----+---------------
  9p  |    3    4    2    total : 9

variable D moyenne    :     0.889 (somme des valeurs :    8)
           variance   :     0.543 (somme des carrs  :   12)
           cart-type :     0.737
           cdv        :    82.916 % (coefficient de variation)
```

```
 Matrice de corrlation

        S          P          M          D

  S     1.0000
  P     0.9608     1.0000
  M     0.9355     0.9934     1.0000
  D     0.0000    -0.2509    -0.2828     1.0000

 -- fin des calculs
```

On pourra comparer avec les résultats pour $n = 6$ :

```
 Etude de 2 ds avec chacun 6 faces
 ... On passe en revue les 36 cas    possibles.

 tude de la variable : S = somme des valeurs

  S  |    1    2    3    4    5    6
-----+------------------------------
  1  |    2    3    4    5    6    7
  2  |    3    4    5    6    7    8
  3  |    4    5    6    7    8    9
  4  |    5    6    7    8    9   10
  5  |    6    7    8    9   10   11
  6  |    7    8    9   10   11   12

  S  |    2    3    4    5    6    7    8    9   10   11   12
-----+-------------------------------------------------------
36p  |    1    2    3    4    5    6    5    4    3    2    1    total : 36

 variable S moyenne    :     7.000 (somme des valeurs :   252)
            variance   :     5.833 (somme des carrs  :  1974)
            cart-type :     2.415
            cdv        :    34.503 % (coefficient de variation)
```

```
tude de la variable : P = produits des valeurs

   P  |    1    2    3    4    5    6
 -----+-------------------------------
   1  |    1    2    3    4    5    6
   2  |    2    4    6    8   10   12
   3  |    3    6    9   12   15   18
   4  |    4    8   12   16   20   24
   5  |    5   10   15   20   25   30
   6  |    6   12   18   24   30   36

   P  |    1    2    3    4    5    6    8    9   10  ...
 -----+---------------------------------------------
 36p  |    1    2    2    3    2    4    2    1    2  ...


     (suite)      12   15   16   18   20   24   25   30   36
                 ---------------------------------------------
                   4    2    1    2    2    2    1    2    1    total : 36

variable P moyenne    :    12.250 (somme des valeurs :    441)
           variance   :    79.965 (somme des carrs  :   8281)
           cart-type :     8.942
           cdv        :    72.999 % (coefficient de variation)

tude de la variable : M = somme*produit des valeurs

   M  |    1    2    3    4    5    6
 -----+-------------------------------
   1  |    2    6   12   20   30   42
   2  |    6   16   30   48   70   96
   3  |   12   30   54   84  120  162
   4  |   20   48   84  128  180  240
   5  |   30   70  120  180  250  330
   6  |   42   96  162  240  330  432
```

```
  M  |    2    6   12   16   20   30   42   48   54   70 ...
-----+----------------------------------------------------
36p  |    1    2    2    1    2    4    2    2    1    2 ...


   (suite)   84   96  120  128  162  180  240  250  330  432
            --------------------------------------------------
              2    2    2    1    2    2    2    1    2    1    total : 36

variable M moyenne    :   106.167 (somme des valeurs :   3822)
           variance   : 11034.528 (somme des carrs  : 803012)
           cart-type :   105.045
           cdv        :    98.944 % (coefficient de variation)

tude de la variable : D = valeur absolue de la diffrence

  D  |    1    2    3    4    5    6
-----+--------------------------------
  1  |    0    1    2    3    4    5
  2  |    1    0    1    2    3    4
  3  |    2    1    0    1    2    3
  4  |    3    2    1    0    1    2
  5  |    4    3    2    1    0    1
  6  |    5    4    3    2    1    0

  D  |    0    1    2    3    4    5
-----+--------------------------------
36p  |    6   10    8    6    4    2    total : 36


variable D moyenne    :     1.944 (somme des valeurs :     70)
           variance   :     2.052 (somme des carrs  :    210)
           cart-type :     1.433
           cdv        :    73.679 % (coefficient de variation)
```

```
Matrice de corrlation

      S         P         M         D
 S    1.0000
 P    0.9453    1.0000
 M    0.9120    0.9889    1.0000
 D    0.0000   -0.2808   -0.2864    1.0000

-- fin des calculs
```

# 3. Lois et valeurs comme $V < m$ pour $\mathcal{B}(n,p)$

### Enoncé

Un élève prétend avoir calculé $m_X = 3.123456$, $V_X = 1.654321$. Est-ce possible ?

Un autre élève prétend avoir trouvé $m_X = 3.123456$, $M_{X^2} = 1.654321$. Est-ce possible ?

Un troisième enfin prétend avoir $V_X = 3.123456$ et $m_X = 1.654321$. Est-ce possible ?

Trouvez deux façons de démontrer que $V_X$ est toujours positif ou nul.

### Solution

Les valeurs $m_X = 3.123456$ et $V_X = 1.654321$, sans aucune autre information, sont possibles. Par contre, $m_X = 3.123456$ et $M_{X^2} = 1.654321$ est impossible car (voir plus bas) $V(X) = M_{X^2} - m_X^2 \geq 0$ et donc on a toujours $M_{X^2} \geq m_X^2$ ce qui n'est pas le cas ici. Enfin, $V_X = 3.123456$ et $m_X = 1.654321$ sans aucune autre information, est possible. Toutefois, si on sait que $X$ est une loi binomiale, c'est impossible car pour une telle loi $V = np(1-p)$ est forcément inférieur à $m = np$ puisque $p$ (et donc $1-p$) est inférieur à 1.

# 4. Loi de *Bernoulli* généralisée

### Enoncé

Soit $\mathsf{U}=b(x,y,p)$ la loi de Bernoulli généralisée, qui prend les valeurs $y$ et $x$ avec les probabilités respectives $p$ et $1-p$ avec $x \leq y$. Calculer directement $m(\mathsf{U})$ et $V(\mathsf{U})$. En remarquant que $\mathsf{U}=a.\mathsf{T}+b$ où $\mathsf{T}=b(p)$, calculer $a$ et $b$ puis retrouver les valeurs de $m(\mathsf{U})$ et $V(\mathsf{U})$.

### Solution

Le calcul direct donne : $m(\mathsf{U}) = (1-p)x + py = x + (y-x)p$ ; de même $V(\mathsf{U}) = (1-p)x^2 + py^2 - m(\mathsf{U})^2$ soit $V(\mathsf{U}) = (y-x)^2p(1-p)$. On peut aussi écrire $\mathsf{U} = (y-x)\mathsf{T} + x$ d'où, comme $m(aX+b) = a.m(X) + b$ et $V(aX+b) = a^2V(X)$, $m(\mathsf{U}) = (y-x)m(\mathsf{T}) + x = (y-x)p + x$ et $V(\mathsf{U}) = (y-x)^2V(\mathsf{T}) = (y-x)^2p(1-p)$.

# 5. Centrage et Réduction

### Enoncé

Soit $X$ une v.a. ; construire une v.a. $Y$ liée linéairement à $X$ telle que $m(Y) = 0$. On la nomme la variable centrée issue de $X$. Construire une v.a. $Z$ liée linéairement à $X$ telle que $\sigma(Z) = 1$. On la nomme la variable réduite issue de $X$. Construire une v.a. $T$ liée linéairement à $X$ telle que $m(T) = 0$ et $\sigma(T) = 1$. On nomme $Z$ la variable centrée réduite issué de $X$. Comparer $T$ avec la variable $X - m(X)/\sigma(X)$ et à la variable $X/\sigma(X) - m(X)$.

### Solution

Posant $Y = aX + b$, la condition $m(Y) = 0$ impose $a = 1$ et $b = -m_X$. De même pour $Z = cX + d$ où la condition $\sigma(Z) = 1$ impose $d = 0$ et $c = 1/\sigma$. Si la variable $T = eX + f$ vérifie $m(T) = 0$ et $\sigma(T) = 1$, alors $e = 1/\sigma$ et $f = -m$ et $T$ est donc la variable $(X-m)/\sigma$. Par contre $m(X - m(X)/\sigma(X)) = m(X)(1-1/\sigma)$ ce qui ne donne rien de remarquable, pas plus que $m(X/\sigma(X) - m(X)) = m(X)(1/\sigma - 1)$.

# 6. Qu'est-ce qu'une moyenne ?

### Enoncé

On appelle moyenne arithmétique de deux valeurs $x$ et $y$ la quantité $(x+y)/2$, moyenne géométrique la quantité $\sqrt{x * y}$, moyenne quadratique $\sqrt{(x^2+y^2)/2}$, moyenne harmonique $2/((1/x)+(1/y))$.

Généraliser à $n$ valeurs $x_1, x_2 ... x_n$ plutôt que $x$ et $y$. Montrer que les différentes fonction-moyennes vérifient les propriétés suivantes :

$$min\{x_i\} \leq moy((x_i)) \leq max\{x_i\}$$

$$\forall i x_i = c \Rightarrow moy((x_i)) = c$$

$moy((x_i))$ est invariante par permutation des $x_i$

Comparer $ma$, $mg$, $mq$ et $mh$ pour $x = 2$, $y = 8$. Et dans le cas général ?

Soient $\alpha_i$ des réels. Quelle(s) condition(s) doit-on imposer aux $\alpha_i$ pour que, les $m_i$ désignant des fonctions-moyennes, la combinaison linéaire $\sum \alpha_i m_i$ soit aussi une fonction-moyenne ?

### Solution

La moyenne arithmétique de $n$ valeurs $x_i$ pour $i$ de 1 à $n$ est $ma((x_i)) = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Leur moyenne géométrique est $mg((x_i)) = \sqrt[n]{\prod_{i=1}^{n} x_i}$.

Leur moyenne quadratique $mq((x_i)) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$.

Leur moyenne harmonique $mh((x_i)) = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$.

Si pour tout $i$, $x_i = c > 0$, alors $ma((x_i)) = n.c/n = c$.

De même $mg((x_i)) = \sqrt[n]{c^n} = c$, $mq((x_i)) = \sqrt{nc^2/n} = c$, $mh((x_i)) = n/(n/c) = c$.

Puisque $min\{x_i\} \leq x_i \leq max\{x_i\}$, en sommant terme à terme puis en divisant terme à terme par $n$, on déduit $min\{x_i\} \leq ma((x_i)) \leq max\{x_i\}$. Pour les autres moyennes, un calcul analogue aboutit aux mêmes conclusions.

L'addition étant commutative, $ma$ n'est pas sensible à une permutation des $x_i$. Même remarque pour les autres moyennes.

Pour $x = 2$, $y = 8$, $ma(x, y) = (8 + 2)/2 = 10/2 = 5$.

$mg(x, y) = \sqrt{8 + 2} = \sqrt{16} = 4$ $mh(x, y) = 1/((1/2) + (1/8)) = 8/5 = 1.6$ $mq(x, y = \sqrt{(2^2 + 8^2)/2} = \sqrt{34} = 5.83$

Donc $mh < mg < ma < mq$.

# Annexe

```
#########################################################
#                                                       #
# deuxdes.tcl : programme de calcul des v.a.          #
#               S,P,M,D et leur caractristiques       #
#               pour deux ds  n faces                #
#                                                       #
#########################################################
#
#             auteur : Gilles HUNAULT
#             email  : gilles.hunault@univ-angers.fr
#             web    : http://www.info.univ-angers.fr/pub/gh/gh.html
#

proc copies { a b } {

 # renvoie b copies de a

 set c ""
 set i 1
 while { $i <= $b } {
   append c $a
   set i [incr i ]
 } ; # fin tant que sur i

 return $c

} ; # fin de proc copies

proc mve { ns } {

 ## affiche les valeurs puis moyenne, variance, cart-type

 global n sdv nom

 puts "\n tude de la variable : $nom($ns,1) = $nom($ns,2)\n"
```

```
# 1. affichage en produit cartsien

# 1.1 affichage de la ligne de label

set i 1
set  ls "    $nom($ns,1)  |"
while { $i <= $n } {
  set f [format "%5d" $i]
  append ls "$f"
  set i [incr i ]
} ; # fin tant que sur i
puts "$ls"
puts "  [copies "-" 5]+[copies "-" 32]"

# 1.1 affichage des lignes de valeurs

set i 1
set v 0
while { $i <= $n } {
  set  j 1
  set  ls ""
  set  f [format "%5d" $i]
  append ls "$f  |"
  while { $j <= $n } {
      set v [incr v]
      set f [format "%5d" [expr int($sdv($v,$ns))] ]
      append ls "$f"
      set j  [incr j]
  } ; # fin tant que sur j
  puts "$ls"
  set i [incr i ]
} ; # fin tant que sur i
```

```
# 2. affichage rsum par valeur

# 2.1 initialisation des effectifs via le max en nom(.,3)

set im  $nom($ns,3)
set i 0
while { $i <= $im } {
  set eff($i) 0
  set i [incr i ]
} ; # fin tant que sur i

# 2.2.1  remplissage des effectifs

set nc  [expr $n*$n]
set i   1
while { $i <= $nc } {
  set es [expr int($sdv($i,$ns))]
  set eff($es) [expr 1 + $eff($es)]
  set i [incr i ]
} ; # fin tant que sur i

# 2.2.2 comptage des valeurs

set  i   0
set  nv  0
set  sde 0
while { $i <= $im } {
  if { $eff($i) > 0 } {
     set nv  [incr nv]
     set sde [expr $sde + $eff($i)]
  } ; # fin de si
  set i [incr i ]
} ; # fin tant que sur i
```

```
# 2.3 affichage de ces valeurs

# 2.3.1 affichage de la ligne de label

set i  0
set ls "\n    $nom($ns,1)  |"
while { $i <= $im } {
  if { $eff($i) > 0 } {
     set f [format "%5d" $i]
     append ls "$f"
  } ; # fin de si
  set i [incr i ]
} ; # fin tant que sur i
puts "$ls"
puts "  [copies "-" 5]+[copies "-" [expr 2+5*$nv]]"

# 2.3.2 affichage des ligne des effectifs

set i  0
set s  0.0
set sc 0.0
set ls "[format "%4d" $nc]p  |"
while { $i <= $im } {
  set  effi $eff($i)
  if { $effi > 0 } {
     set f  [format "%5d" $effi ]
     set s  [expr $s   + 1.0*$effi*$i]
     set sc [expr $sc  + 1.0*$effi*$i*$i]
     append ls "$f"
  } ; # fin de si
  set i [incr i ]
} ; # fin tant que sur i
puts "$ls    total : $sde"
```

```
 # 2.4 indicateurs numriques

 set m [expr $s/$nc]
 set v [expr ($sc/$nc)-$m*$m]
 set e [expr sqrt($v)]
 set c [expr 100*$e/$m]

 set fm [format "%9.3f" $m]
 set fv [format "%9.3f" $v]
 set fe [format "%9.3f" $e]
 set fc [format "%9.3f" $c]

 puts "\n variable $nom($ns,1) moyenne    : \
      $fm   (somme des valeurs :            \
      [format "%10.0f" $s])"

 puts "             variance   : \
      $fv   (somme des carrs des valeurs : \
      [format "%10.0f" $sc])"

 puts "             cart-type : $fe"

 puts "             cdv        : $fc \
     \% (coefficient de variation)"

} ; # fin de proc mve

proc corr { ia ib } {

  ## calcule le coefficient de corrélation
  ## linéaire entre deux séries de n valeurs mises
  ## dans le tableau (global) sdv
  ## en colonnes ia et ib

  global n sdv nom
```

```
  if { $ia == $ib } { set cor 1 } else {

     set cor 0
     set nc [expr $n*$n]
     set v   1
     set sa  0
     set sca 0
     set sb  0
     set scb 0
     set sab 0

     while { $v <= $nc } {
       set  xa  $sdv($v,$ia)
       set  xb  $sdv($v,$ib)
       set  xab $xa*$xb
       set  sa  [expr $sa  + $xa]
       set  sca [expr $sca + $xa*$xa]
       set  sb  [expr $sb  + $xb]
       set  scb [expr $scb + $xb*$xb]
       set  sab [expr $sab + $xab]
       set  v   [incr v ]
     } ; # fin tant que sur i

     set ma  [expr $sa/$nc]
     set mca [expr $sca/$nc]
     set mb  [expr $sb/$nc]
     set mcb [expr $scb/$nc]
     set mab [expr $sab/$nc]
     set va  [expr $mca-$ma*$ma]
     set vb  [expr $mcb-$mb*$mb]
     set cov [expr $mab-$ma*$mb]
     set cor [expr $cov/(sqrt($va*$vb))]

  } ; # fin si

  return $cor

} ; # fin de proc corr
```

```
#######################################################
##                                                   ##
##  Programme principal                              ##
##                                                   ##
##                                                   ##
#######################################################

 set n 6 ; #### par paramtre : set n $argv

 puts " Etude de 2 ds avec chacun $n faces "

 set nom(1,1) "S" ;  set nom(1,2) "somme des valeurs"
 set nom(1,3) [expr $n+$n]
 set nom(2,1) "P" ;  set nom(2,2) "produits des valeurs"
 set nom(2,3) [expr $n*$n]
 set nom(3,1) "M" ;  set nom(3,2) "somme*produit des valeurs "
 set nom(3,3) [expr ($n+$n)*($n*$n)]
 set nom(4,1) "D" ;  set nom(4,2) "valeur absolue de la diffrence"
 set nom(4,3) [expr $n-1]

 set i 1
 set v 0
 while { $i <= $n } {
   set  j 1
   while { $j <= $n } {
       set v [incr v]
       set sdv($v,1) [expr 1.0*($i + $j)]
       set sdv($v,2) [expr 1.0*($i * $j)]
       set sdv($v,3) [expr 1.0*(($i * $j)*($i + $j))]
       set sdv($v,4) [expr 1.0*(abs($i - $j))]
       set j  [incr j]
   } ; # fin tant que sur j
   set i [incr i ]
 } ; # fin tant que sur i

 puts " ... On passe en revue les [expr $n*$n] cas    possibles."
```

```
# affichage par variable

set nx 4
set ix 1
while { $ix <= $nx } {
   set tmp [mve $ix]
   set ix [incr ix ]
} ; # fin tant que sur ix

# affichage par couple de variables

puts "\n Matrice de corrlation \n"
set ix 1
while { $ix <= $nx } {
   set ls "    "
   set jx 1
   while { $jx <= $ix } {
       set tc [corr $ix $jx]
       set fc [format "%10.4f" $tc]
       append ls "$fc"
       set jx [incr jx ]
   } ; # fin tant que sur jx
   puts $ls
   set ix [incr ix ]
} ; # fin tant que sur ix

puts "\n -- fin des calculs "
```