Chapitre IV COUPLES DE VARIABLES ALEATOIRES,VECTEURS ALEATOIRES

A) COUPLES DE VARIABLES ALEATOIRES.

Un couple de variables aléatoires (ou vecteur aléatoire de dimension 2) est une application d'un espace probabilisé (Ω,T,P) dans **R** ^{2} , noté (X, Y), tel que pour tous x et y dans **R**, $\{\omega \in \Omega$ / $X(\omega)<x$ et $Y(\omega)<y\} \in$ T.
On montre que X et Y sont des variables aléatoires réelles , X est la première composante et Y la seconde.

Exemple: On jette un dé une première fois, le nombre porté par la face supérieure est une variable aléatoire X; on jette à nouveau le dé jusqu'à ce que l'on obtienne un nombre différent du premier jet, le nombre obtenu est une variable aléatoire Y. On obtient ainsi un couple de variables aléatoires (X, Y).
L' image du couple peut être représentée par l'ensemble des couples (i, j) avec i ≠ j et i et j dans {1,..., 6}, le référentiel Ω est infini et peut être représenté par l'ensemble des suites d'entiers à valeurs dans {1,..., 6}.

Comme dans le cas d'une variable aléatoire réelle, on distingue deux catégories de couples de VAR suivant que l'image est discrète ou continue.

1) Couples de variables aléatoires discrètes.

Le couple sera dit discret lorsque son image est un ensemble fini ou dénombrable. La loi de probabilité du couple est la donnée de l'image du couple Im(X,Y) et des différentes probabilités P[X = x et Y = y] lorsque (x, y) est dans Im(X,Y). Dans le cas discret fini, la loi est représentée sous la forme d'un tableau appelé tableau de contingence.

Exemple: tableau de contingence du couple cité en introduction.

X\Y	**1**	**2**	**3**	**4**	**5**	**6**
1	0	1/30	1/30	1/30	1/30	1/30
2	1/30	0	1/30	1/30	1/30	1/30
3	1/30	1/30	0	1/30	1/30	1/30
4	1/30	1/30	1/30	0	1/30	1/30
5	1/30	1/30	1/30	1/30	0	1/30
6	1/30	1/30	1/30	1/30	1/30	0

La **fonction de répartition du couple** est la fonction qui à (a, b) dans $\mathbf{R}^2$ associe le réel P[X <a et Y < b].
Si g est une fonction de deux variables, **l'espérance mathématique** de la fonction u par rapport au couple est le nombre s'il existe égal à:

$$\mathbf{E}(g(X, Y)) = \sum_{(x,y)\in \mathrm{Im}(X,Y)} g(x,y) P[X = x \text{ et } Y = y]$$

En particulier, la moyenne du couple est **E**(XY), la covariance du couple, notée **Cov** (X, Y)= **E**[(X-**E**(X)(Y-**E**(Y)].
E est aussi un opérateur linéaire défini dans l'espace vectoriel des fonctions de deux variables.

Exemple: Dans le cas du couple étudié, on a **E** (XY)= 35/3.

La loi de probabilité de X (resp. Y) est appelée **loi marginale** de X, elle s'obtient par addition des colonnes (resp. des lignes) dans le tableau de contingence.
Dans l'exemple de référence, on obtient:

X\Y	**1**	**2**	**3**	**4**	**5**	**6**	**P[X =k]**
1	0	1/30	1/30	1/30	1/30	1/30	1/6
2	1/30	0	1/30	1/30	1/30	1/30	1/6
3	1/30	1/30	0	1/30	1/30	1/30	1/6
4	1/30	1/30	1/30	0	1/30	1/30	1/6
5	1/30	1/30	1/30	1/30	0	1/30	1/6
6	1/30	1/30	1/30	1/30	1/30	0	1/6
P[Y=i]	1/6	1/6	1/6	1/6	1/6	1/6	

Ainsi la moyenne de X est égale à **E** (X)= $\sum_{x\in \mathrm{Im}\,X} xP[X = x] = \sum_{(x,y)\in \mathrm{Im}\,(X,Y)} xP[X = x \text{ et } Y = y]$ (dans l'exemple précédent 3,5).

La **loi conditionnelle** de X par rapport à Y= y est définie par son image ImX et les différentes valeurs P[X =x / Y =y]= P[X =x et Y =y] /P[Y =y] , lorsque x est dans ImX.

2) Couples de variables aléatoires continus.

Le couple sera dit continu si l'image du couple est une partie D quarrable de $\mathbf{R}^2$. On suppose dans ce cas qu'il existe une fonction f de deux variables positive intégrable sur $\mathbf{R}^2$, d'intégrale égale à 1, appelée **densité de probabilité** du couple telle que pour tout ensemble Δ quarrable de $\mathbf{R}^2$**,** on ait:

$$P[(X, Y)\in \Delta]= \iint_{\Delta} f(t,u)\,dtdu$$

La **fonction de répartition du couple** est la fonction qui à (a, b) dans $\mathbf{R}^2$ associe le réel $F(a, b)=P[X <a \text{ et } Y <b]= \iint_{t<a \text{ et } u<b} f(t,u)dtdu$

On rappelle que si F est C_2, $\dfrac{\partial^2 F}{\partial x \partial y} = f$.

Si u est une fonction de deux variables, **l'espérance mathématique** de la fonction g par rapport au couple est le nombre s'il existe égal à:

$$\mathbf{E}(g(X, Y)) = \iint_{R^2} f(t,u)\, g(t,u)\, dtdu$$

En particulier, la moyenne du couple est **E**(XY), la covariance du couple, notée **Cov**(X, Y)= **E** ((X- **E** (X)(Y- **E** (Y)).

E est aussi un **opérateur linéaire** défini dans l'espace vectoriel des fonctions de deux variables.

La loi de probabilité de X (resp. Y) est appelée **loi marginale** de X, elle s'obtient à partir de sa fonction de répartition que l'on notera G.

$$\text{Pour x réel, } P[X< x] = \iint_{t<x \text{ et } u \text{ réel}} f(t,u)dtdu$$

On obtient la densité de probabilité h de X par dérivation de la fonction de répartition lorsque f est continue soit $h(x)= \int_{-\infty}^{+\infty} f(x, u)du$.

Si [a b] est un intervalle de **R**, la **loi conditionnelle** de X par rapport à $Y \in [a\ b]$ est définie par sa fonction de répartition.

Si x est dans **R**, $P[X <x \,/\, Y \in [a\ b]\,] = \dfrac{\iint_{t<x \text{ et } a<u<b} f(t, u)dtdu}{\iint_{t \text{ réel et } a<u<b} f(t,u)dtdu}$. La densité s'obtient par dérivation.

3) Indépendance de deux variables aléatoires.

Si X et Y est un couple de variables aléatoires, on dira que X et Y sont indépendantes si et seulement si la fonction de répartition du couple est égale au produit des fonctions de répartitions marginales.
Dans le cas discret la condition précédente est équivalente à pour tous x et y dans Im(X,Y), P[X= x et Y= y]=P[X= x]P[Y= y].

Dans le cas continu la condition précédente est équivalente à la densité du couple est égale au produit des densités marginales.

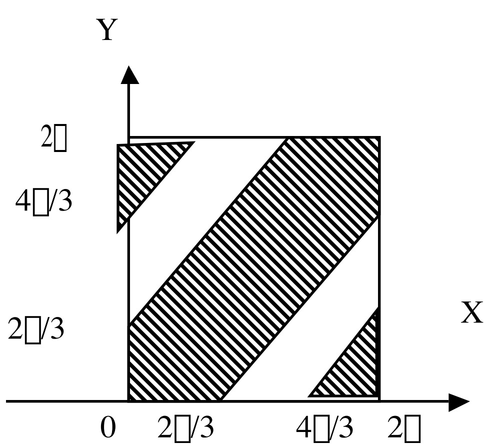
On montre le théorème suivant : si **X et Y sont indépendantes, E(XY)=E(X)E(Y),** théorème sans réciprocité comme le montre l'exemple suivant :

Y \ X	-1	0	1
-1	0	1/4	0
0	1/4	0	1/4
1	0	1/4	0

Dans cet exemple, E(XY)=E(X)=E(Y)=0 et X et Y ne sont pas indépendantes.

Exemple 1 : Le paradoxe de Bertrand (cf. chapitre 2).

On propose une nouvelle modélisation de la corde. Une origine et un sens de rotation étant choisis, les deux extrémités de la corde, X et Y sont repérées par leur angles . On suppose que X (resp. Y) est une variable aléatoire continue uniforme sur l'intervalle $[0\ 2\pi[$ et que X et Y sont indépendantes.
La longueur de la corde est inférieure au côté du triangle correspond à l'événement « $|X-Y| \leq 2\pi/3$ ou $|X-Y| \geq 4\pi/3$ ».



La probabilité cherchée est le rapport de l'aire hachurée sur l'aire du carré soit 6/9= 2/3, on retrouve la même probabilité correspondant à la au premier cas;

Exemple 2 : L'aiguille de Buffon.

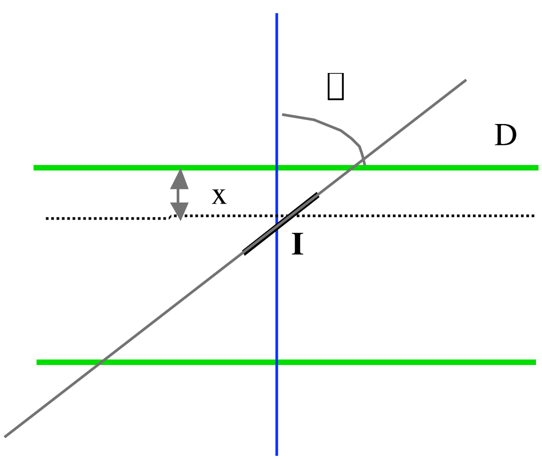
Le jeu de l'aiguille est un exemple de technique permettant de résoudre un problème numérique à l'aide d'un procédé statistique que l'on appelle aujourd'hui une méthode de MONTE CARLO.

On suppose tracées sur un plan horizontal des parallèles équidistantes à la distance 2a. On laisse tomber une aiguille de longueur a sur ce plan et on cherche la probabilité pour que l'aiguille coupe une des parallèles.

1) Modélisation du phénomène et calcul de la probabilité.

Les parallèles étant toutes équivalentes, on peut réduire l'étude à deux parallèles consécutives. Comme il y a symétrie par rapport à la parallèle équidistante, et homogénéité, suivant la direction des parallèles, on peut considérer que le problème s'énonce de la façon suivante:

La position de l'aiguille est déterminée à partir de la distance X du milieu I de l'aiguille à la parallèle la plus proche et de l'angle fait par l'aiguille avec la direction perpendiculaire aux parallèles.
On suppose que X suit une loi continue uniforme sur l'intervalle [0 a] et que □ une loi continue uniforme sur l'intervalle[0 □/2] et que les deux variables aléatoires X et □ sont indépendantes.



Ainsi, la densité du couple (X , □) est égale à 2/□a si x est dans l'intervalle [0 a] et □ dans l'intervalle [0 □/2] et 0 sinon.
L'aiguille coupe la parallèle D si arcos (□) >2X.
La probabilité cherchée est obtenue en intégrant la fonction densité dans le domaine A défini par :

0< x <a , 0<□<□/2 et arcos(□)>2x.

Soit $\iiint_{\square} \frac{2}{a\square} dxd\square = \int_{0}^{\frac{\square}{2}} \frac{a}{2} \cos\square \, d\square = 1/\square.$

2) Calcul du nombre □ par la méthode de Monte Carlo.

Compte tenu de la loi des grands nombres, on peut dire que si l'on jette l'aiguille N fois , et que l'on observe N' fois l'événement l'aiguille coupe une parallèle, la fréquence d'observation de l'événement, N'/N est d'autant plus proche de 1/□ que N est grand. Cette méthode permet de calculer la valeur de □ avec une précision dépendant du nombre N d'expériences.
Il faut noter que le jeu peut se prolonger jusqu'à ce que l'imperfection de l'aiguille et des parallèles rend illusoire de continuer

4) Somme de deux variables aléatoires.

Si X et Y sont des variables aléatoires et si φ est une fonction continue de deux variables alors si on pose $Z = \varphi(X,Y)$, Z est aussi une variable aléatoire continue si X et Y sont continues et discrète si X et Y sont discrètes. A titre d'exemple on va étudier la somme de deux variables aléatoires indépendantes.

A) Cas discret :

Si $Z = X + Y$ alors $ImZ = \{ x + y / x \in ImX \text{ et } y \in ImY \}$ et . si $z \in ImZ$,

$$P[Z = z] = P[X + Y = z] = \sum_{x+y=z} P[X = x]P[Y = y].$$

Si X suit une loi B(n, p) et Y une loi B(n', p), et si X et Y sont indépendantes, (voir chapitre III) alors

$ImZ = \{ 0,1,\dots, n+n'\}$ et si $k \in ImZ$,

$$P[Z = k] = \sum_{i=0}^{k} P[X = i \text{ et } Y = k-i] = \sum_{i=0}^{k} P[X = i]P[Y = k-i] =$$

$$\sum_{i=0}^{k} C_n^i p^i (1-p)^{n-i} C_{n'}^{k-i} p^{k-i} (1-p)^{n'-k+i} = \sum_{i=0}^{k} C_n^i C_{n'}^{k-i} p^k (1-p)^{n+n'-k}$$

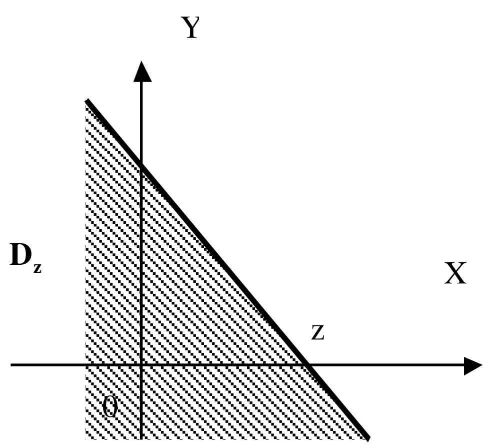
$$= C_{n+n'}^k p^k (1-p)^{n+n'-k}$$

ainsi Z suit une B(n+n', p).

B) Cas continu

Si X est de densité f et Y de densité g alors la fonction de répartition de Z,

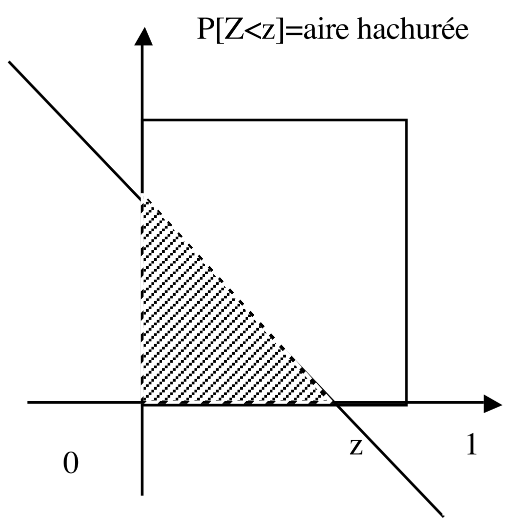
$P[Z < z] = P[X + Y < z] = \iint_{D_z} f(u)g(v)\,du\,dv$ ou D_Z représente le domaine $\{ (x, y) \in \mathbf{R}^2 / x+y < z \}$.

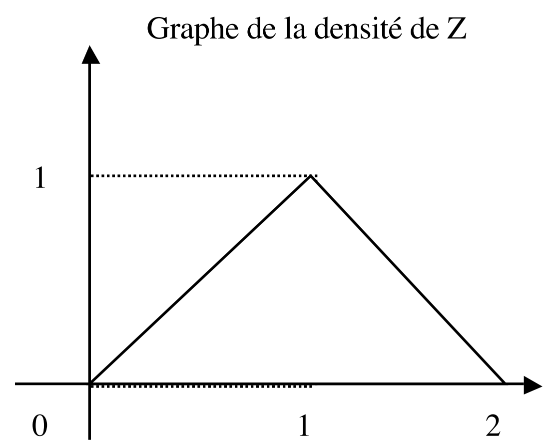


Exemple : On suppose que X et Y sont indépendantes et suivent une loi uniforme sur l'intervalle [0, 1] quelle est la loi de Z = X +Y ?

$$P[Z < z] = \iint_{D_z} f(u)g(v)\,du\,dv = \begin{cases} 0 \text{ si } z \leq 0 \\ \dfrac{z^2}{2} \text{ si } 0 < z < 1 \\ 1 - \dfrac{(2-z)^2}{2} \text{ si } 1 \leq z < 2 \\ 1 \text{ si } z \geq 2 \end{cases}$$

La densité est la dérivée soit, $f(z) = \begin{cases} 0 \text{ si } z \leq 0 \\ z \text{ si } 0 < z < 1 \\ z-2 \text{ si } 1 \leq z < 2 \\ 0 \text{ si } z \geq 2 \end{cases}$.





On montre le résultat suivant concernant la somme de deux variables aléatoires:

THEOREME : Si X et Y sont deux variables aléatoires alors **E**[X+Y] =**E**[X] +**E**[Y] et **V**[X+Y] =**V**[X] +**V**[Y] +2**Cov**(X,Y) et ainsi , lorsque les variables sont indépendantes, **V**[X+Y] =**V**[X] +**V**[Y]

Ce théorème se généralise à une somme finie de variables aléatoires.

THEOREME : Si X_1 , X_2,..., X_n sont n variables aléatoires alors $\mathbf{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E(X_i)$ et si de plus les variables sont indépendantes, $\mathbf{V}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} V(X_i)$.

B) LIAISON ENTRE DEUX VARIABLES ALEATOIRES.

Si X et Y sont deux variables aléatoires, on définit deux nombres caractérisant la liaison entre les variables : la covariance notée **Cov**(X,Y) = **E**[(X-**E**(X))(Y-**E**(Y)] et le coefficient de corrélation noté $\rho_{XY} = \dfrac{Cov(X, Y)}{\sigma_X \sigma_Y}$.

L'opérateur **E** étant linéaire, on montre que **Cov**(X,Y) = **E**(XY)-**E**(X)**E**(Y) et par conséquent, si X et Y sont indépendantes, la covariance et le coefficient de corrélation sont nuls, propriété sans réciprocité.

L'opérateur covariance est une forme bilinéaire symétrique , sa forme quadratique associée est par conséquent la variance.
Si X et Y sont eux VAR définies sur le même espace probabilisé et si λ est un réel variable , $\mathbf{V}(X+\lambda Y) = \lambda^2\mathbf{V}(Y)+2\lambda\mathbf{Cov}(X, Y)+\mathbf{V}(X)$ est un trinôme du second degré en λ.
Comme une variance est toujours positive, le discriminant est négatif ou nul et on obtient le théorème :

THEOREME : Si X et Y sont deux variables aléatoires, alors :

$$\mathbf{Cov}(X,Y)^2 \leq \mathbf{V}(X)\mathbf{V}(Y)$$

c'est l'inégalité de Cauchy -Schwarz dans les espaces préhilbertiens réels et par conséquent le coefficient de corrélation est un nombre de l'intervalle [-1,1].

Lorsque il y a égalité, le trinôme du second degré a une racine double $\lambda_0 = -\dfrac{Cov(X, Y)}{V(Y)}$ et ainsi $V(X+\lambda_0 Y) = 0$

Dans ce cas le coefficient de corrélation est $\rho_{XY} = \pm 1$ et $\lambda_0 = -\dfrac{\sigma_X \rho_{XY}}{\sigma_Y}$

L'inégalité de Bienaymé - Tchebychev, permet d'affirmer que $P[X+\lambda_0 Y = E(X+\lambda_0 Y)]=1$.

On dit que X et Y sont linéairement liées presque partout c'est à dire que il existe des réel a et A, tels que P[X= aY+A] = 1.

Plus précisément, on obtient:

- Si ρ_{XY} =1 P[X- $\frac{\sigma_X}{\sigma_Y}$ Y= **E**(X)- $\frac{\sigma_X}{\sigma_Y}$ **E**(Y)] =1 on dit que X et Y sont corrélées.

-Si ρ_{XY} =-1 P[X+ $\frac{\sigma_X}{\sigma_Y}$ Y= **E**(X)+ $\frac{\sigma_X}{\sigma_Y}$ **E**(Y)]] =1 on dit que X et Y sont anticorrélées.

Exemple 1 : ETUDE D'UN NUAGE DE POINTS

On considère sur une même population deux variables quantitatives X et Y, pour lesquelles on dispose de n observations, résumées dans le tableau ci-dessous :

X	x_1	x_2	x_3	…	…	x_n
Y	y_1	y_2	y_3	…	…	y_n

Par exemple, X désigne la température et Y la consommation de fuel chaque jours pendant une certaine période.
Pour avoir une idée de la relation entre les deux variables, on suppose qu'elles sont équiréparties c'est à dire que P[X =x et Y =y]=1/n.

On obtient :

La moyenne **E**(X)= $\frac{1}{n}\sum_{i=1}^{n} x_i$, la variance **V**[X] = $\frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))^2 = \frac{1}{n}[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2]$;

la moyenne **E**(Y)= $\frac{1}{n}\sum_{i=1}^{n} y_i$, la variance **V**[Y] = $\frac{1}{n}[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2]$;

la covariance **Cov**(X,Y)= $\frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))(y_i - E(Y)) = \frac{1}{n}[\sum_{i=1}^{n} x_i y_i - \frac{1}{n}(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)]$;

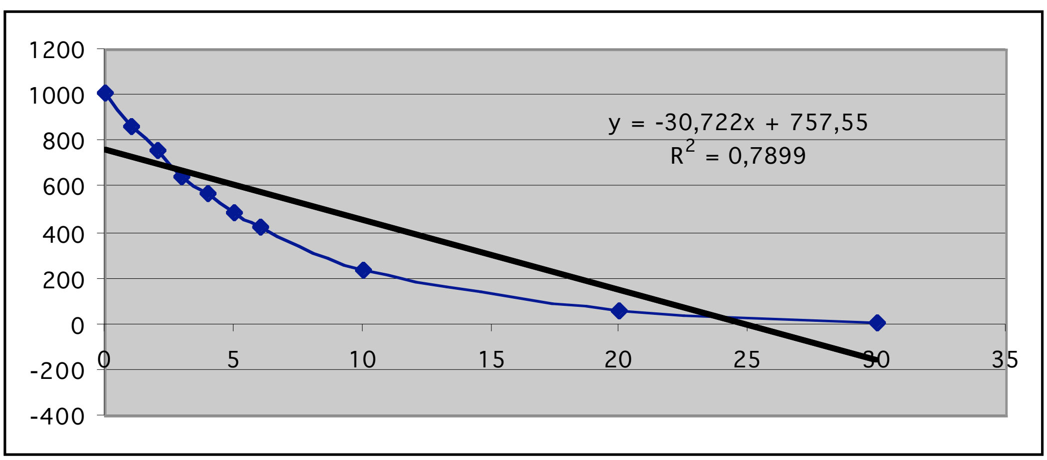
le coefficient de corrélation linéaire $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$.

Si le coefficient de corrélation linéaire est égal à ±1, les variables X et Y sont linéairement liées et par conséquent σ_Y(X-E(X)) = ± σ_X(Y-E(Y))= $\rho_{X,Y}$ σ_X(Y-E(Y)).
On peut noter que la droite passa par le point moyen (E(X),E(Y)) et que son équation s'écrit Y-E(Y) = $\frac{Cov(X,Y)}{V(X)}$(X-E(X)) ou X-E(X) = $\frac{Cov(X,Y)}{V(Y)}$(Y-E(Y))

Lorsque le coefficient de corrélation est proche de ±1(en pratique supérieure en valeur absolue à 0,7avec n>15), les équations ci-dessus ne coïncident pas et on obtient deux droites qui sont d'autant plus proches que $\rho_{X,Y}$ est proche de ±1.
La régression linéaire n'est qu'une manière commode d'obtenir une droite à partir de n relevés expérimentaux , les calculatrices de poches disposent souvent de programmes automatiques de la droite de régression de y en x et du coefficient de corrélation.
La VAR ε= Y-aX-b est l'erreur que l'on fait lorsque on remplace Y par la VAR obtenue à partir de la régression linéaire . On montre facilement que cette variable est centrée et de variance $\mathbf{V}(\varepsilon)=\mathbf{V}(Y)(1-\rho^2)$ et ainsi, $\rho^2 = \dfrac{V(Y) - V(\varepsilon)}{V(Y)}$.
La variance $\mathbf{V}(\varepsilon)$ est une variance résiduelle qui n'est pas expliquée par la droite de régression, on obtient $V(Y)= \mathbf{V}(Y)(1-\rho^2) + V(aX+b)= \mathbf{V}(Y)(1-\rho^2) + \rho^2 V(Y)$ et ainsi, la variance de Y est la somme de la variance résiduelle et de la variance expliquée par la droite de régression.

Application: L'altitude est le facteur essentiel qui conditionne la pression atmosphérique. On a mesuré la pression X en hPA suivant l'altitude Y en km, le modèle étant exponentiel, on a calculé LN(Y) et obtenu le tableau ci-dessous :

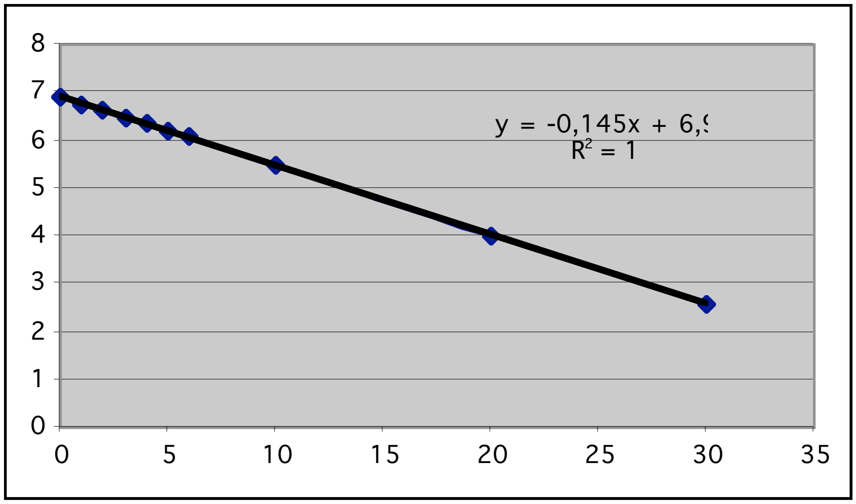
X	0	1	2	3	4	5	6	10	20	30
Y	1013	865	760	650	570	490	430	240	56	13
LN(Y)	6,9206	6,7627	6,6333	6,476	6,3456	6,1944	6,0637	5,4806	4,0253	2,565



L'étude de la régression de Y en fonction de X donne un coefficient de corrélation égal à R=-0,888 et une représentation graphique qui n'est pas satisfaisante puisque la variance résiduelle est 23281.

L'étude de la régression de LN(Y) en fonction de X donne un coefficient de corrélation égal à R=-0,9999 et de variance résiduelle 0 donne une représentation graphique bien meilleure.

Ainsi , on peut estimer que l'on a la relation Y= exp(-0,145X+6,21) et on peut vérifier que la pression est divisée par dix lorsque on monte de 16 km.



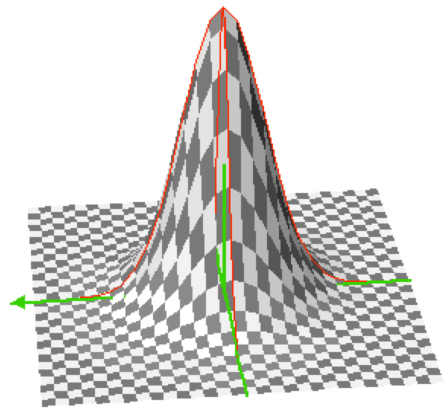
Exemple 2 : Vecteur Gaussien en dimension deux.

Un couple de VAR (X,Y) est un vecteur Gaussien lorsque sa densité de probabilité est :

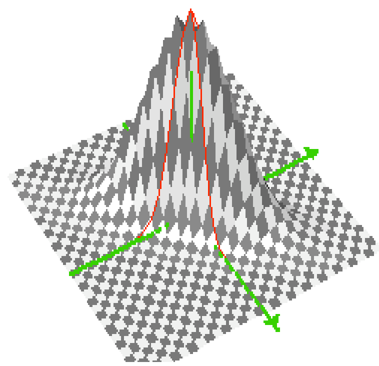
$$f(x\,,\,y)= \frac{1}{2\pi\sigma\sigma'\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}[(\frac{x-m}{\sigma})^2-2\rho\frac{(x-m)(y-m')}{\sigma\sigma'}+(\frac{y-m}{\sigma'})^2]\right]$$

ρ est un réel de]-1,1[, σ et σ' des réels strictement positifs.

On montre que ρ est le coefficient de corrélation linéaire de X et Y , que la variable X (resp. Y) suit une loi normale N(m, σ) (resp. Y une N(m', σ')) et que X et Y sont indépendantes si et seulement si $\rho = 0$.



Graphe de la densité du vecteur gaussien lorsque □ = 0



Graphe de la densité du vecteur gaussien lorsque □ = 0,8

Exemple 3 LA LOI MULTINOMIALE.

Une urne contient trois types de boules et la probabilité de voir apparaître une boule d'un type déterminé au cours d'un tirage est notée p_1, p_2 et p_3. On effectue n tirages successifs avec remise.

On notera que p_1 est la proportion de boules ayant le premier type (resp. p_2, p_3) et par conséquent, $p_1 + p_2 + p_3 = 1$.

Si X est la variable aléatoire correspondant au nombre de boules tirées du type 1, si Y est la variable aléatoire correspondant au nombre de boules tirées du type 2, et si Z est la variable aléatoire correspondant au nombre de boules tirées du type 3, la variable Z =n -X +Y.

La loi de probabilité du couple (X, Y) est donnée par les différentes valeurs :

$$P[X = k_1 \text{ et } Y = k_2] = \frac{n!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}$$

Avec k_1, k_2, k_3 entiers naturels tels que $k_1 + k_2 + k_3 = n$.

Il est clair que X suit une $B(n, p_1)$, Y suit une $B(n, p_2)$ et que Z suit une $B(n, p_3)$. On obtient la covariance de X et Y en considérant $\mathbf{V}(X+Y) = \mathbf{V}(X)+\mathbf{V}(Y)+2\mathbf{Cov}(X, Y) = \mathbf{V}(n - Z)= \mathbf{V}(Z)$ et ainsi $\mathbf{Cov}(X,Y)=-np_1p_2$.

Les variables X et Y ne sont pas indépendantes puisque p_1 et p_2 sont non nuls, on dit que le triplet (X, Y, Z) suit une loi multinomiale d'ordre trois (ou trinomiale).
Le cadre général est celui d'une épreuve ayant trois résultats possibles et répétée n fois dans des conditions identiques.
La variable aléatoire $T=\frac{(X-np_1)^2}{np_1}+\frac{(Y-np_2)^2}{np_2}+\frac{(Z-np_3)^2}{np_3}$ est utilisée dans la procédure du test du Khi-deux étudié au chapitre suivant.
On peut en développant les termes carrés montrer que la variable aléatoire T est égale à :

$$T=\frac{1-p_2}{np_3}[X-np_1+\frac{p_1}{1-p_2}(Y-np_2)]^2+\frac{(Y-np_2)^2}{np_2(1-p_2)}.$$

Ainsi, on vérifie que T est la somme de carrés de deux variables aléatoires centrées réduites et indépendantes.
Lorsque n tends vers l'infini(en pratique, n est assez grand) la loi de T suit une loi de Khi-deux ayant deux degrés de liberté.
Ce résultat se généralise à une loi multinomiale d'ordre n et permet ainsi de mettre en place les tests dits de Khi-deux.

EXERCICES DU CHAPITRE IV

Exercice1

Soit X et Y deux variables aléatoires indépendantes de Poisson de paramètre respectivement m et n.
a)Quelle est la loi de Probabilité de Z = X + Y ?
b) Si k est un entier fixé, donner la loi de probabilité de la variable aléatoire X conditionnée à X + Y = k .

Exercice 2

Le nombre de personnes se présentant à un guichet donné d'une poste pendant une période de 15 minutes est une variable aléatoire X de poisson de paramètre m.
La probabilité pour qu'une personne se présentant au guichet demande des timbres est 0,6. Soit Y la variable aléatoire égale au nombre de personnes ayant demandé des timbres durant une période de 15 minutes.

1) Calculer la probabilité de l'événement [X= i et Y = j] lorsque i et j sont des entiers naturels.

2)En déduire la loi de probabilité de Y , **E**(Y) et **V**(Y).

3) X-Y désigne la variable aléatoire égale au nombre de personnes n'ayant pas demandé des timbres durant une période de 15 minutes. Donner la loi de X-Y.

4) Comparer **V**(Y), **V**(X), **V**(X-Y)

5)Les variables Y et X-Y sont elles indépendantes ?

6) Soit Z la VAR correspondant au nombre de personnes se présentant au guichet en une heure et T le nombre de personnes demandant des timbres au guichet en une heure.
Donner les lois de Z et de T.

Exercice 3

Une variable aléatoire discrète X prend les valeurs 1, 2, 3, 6. Pour k = 1, 2, 3, 6, la probabilité de l'événement [X =k] est proportionnelle à k.

1) Donner la loi de X et calculer la moyenne et la variance de X.

2) Soit alors la variable aléatoire $Y = (X-3)^2$. Donner la loi du couple (X,Y) et calculer le coefficient de corrélation entre X et Y.

Exercice 4

Un étudiant subit 3 contrôles pendant l'année scolaire. On suppose que pour i = 1, 2, 3, les variables aléatoires Y_i : "note au i -ème contrôle" sont normales d'espérance m et de variance σ^2 et indépendantes . On suppose de plus que :
$P[Y_i \geq 10] = 0,305$ et $P[Y_i < 16] = 0,994$.
a) Calculer m et σ.
b) Donner la loi de Y , note moyenne des 3 contrôles et calculer $P[Y > 10]$.
c) Soit Z la VAR correspondant à la meilleure des 3 notes. Donner la densité de probabilité de Z et calculer $P[Z \geq 10]$.

Exercice 5

Une urne contient n boules numérotées de 1 à n ($n \geq 3$). On tire *simultanément* 3 boules de cette urne et on note X, Y, M les variables aléatoires respectivement égales au plus petit numéro tiré, au plus grand numéro tiré et au numéro médian (ou intermédiaire).

a) Quelles sont les valeurs possibles de M ? Quand M = m, quelles sont les valeurs possibles du couple (X,Y) ? En déduire la loi de probabilité de M.
b) Déterminer la loi de probabilité conjointe du couple (X,Y).

Exercice 6

On désigne par X et Y les longueurs des deux côtés de l'angle droit d'un triangle rectangle.
Si on suppose que X et Y sont deux variables aléatoires normales centrées réduites et indépendantes, donner la loi de probabilité de la variable Z^2 correspondant à la longueur au carré de l'hypoténuse.
Quelle est la moyenne et la variance de Z^2 ?

Exercice 7

Deux personnes A et B se fixent rendez-vous entre 8 h et 9 h et conviennent de ne pas s'attendre plus de 10 minutes.
On suppose que leurs instants d'arrivée sont indépendants et uniformément répartis entre 8 h et 9 h. Plus précisément, si on note A (resp. B) la variable aléatoire correspondant à

l'instant d'arrivée de A (resp. B) , A suit une loi continue uniforme sur l' intervalle [8, 9] et A et B sont indépendantes.

1) Calculer la probabilité de rencontre des deux personnes.

2) Si x est l'instant d'arrivée de A, calculer la probabilité conditionnelle de rencontre sachant A = x.

Exercice 8

Une urne contient des boules blanches en proportion p, des boules noires en proportion q et des boules vertes en proportion r. (p+q+r=1)

On tire n boules une à une avec remise et on note X, Y et Z les VAR correspondant au nombre de boules blanches, noires et vertes tirées.

1) Quelle est la loi de X ?, la loi de Y?, la loi de Z?

2) Calculer la probabilité de l'événement [X= i et Y = j] lorsque i et j sont des entiers naturels.

En déduire la somme $\sum_{i=0}^{n}\sum_{j=0}^{n-i}$ P [X= i et Y = j]

3) Quelle est la loi de la variable X+Y ?

4) Calculer **Cov**(X, Y) et le coefficient de corrélation linéaire $\rho_{X,Y}$.

Exercice 9

Soient U et V deux variables aléatoires indépendants. On note respectivement F et G leurs fonction de répartition. Soit Y= inf. (U,V) .

1)Déterminer en fonction de F et G la fonction de répartition H de Y.

2)On suppose que U et V sont continues, indépendantes et uniformes sur l'intervalle [0,1] . Donner la fonction de répartition de Y, la densité de probabilité de Y et calculer **E**(Y) et **V**(Y).

3)Soient $X_1, X_2, \ldots, X_n$ n variables aléatoires indépendantes continues et uniformes sur l'intervalle [0,1] .

Donner la fonction de répartition de la variable aléatoire $Y_n = \inf(X_1, X_2, \ldots, X_n)$ et en déduire la densité de probabilité. Calculer $\mathbf{E}(Y_n)$ et $\mathbf{V}(Y_n)$.

Exercice 10

Un examen se déroule sous forme d'un questionnaire à choix multiples (QCM) où on pose 20 questions ; chaque question comporte quatre réponses possibles, dont une et une seule est la bonne; une réponse juste compte 1 point , sinon zéro.
On suppose que le programme de l'examen comporte 100 questions dont on tire aléatoirement les 20 de l'examen.
On considère un candidat ayant appris une proportion p de questions du programme. Ce candidat choisit la stratégie suivante: il répond aux questions dont il connaît la réponse et pour les autres, il choisit au hasard une réponse parmi les quatre proposées.

1) Montrer que la variable aléatoire X correspondant au nombre de questions dont il connaît la réponse suit une loi hypergéométrique dont on donnera les paramètres.

2) On note N la variable aléatoire correspondant à la note obtenue par le candidat. Montrer que $N = X + Y$ ou Y est une variable aléatoire telle que la loi conditionnelle de Y sachant $X = x$ suit une loi binomiale $B(20-x; 0{,}25)$.

Montrer que si $n = 0, \ldots, 20$ la probabilité $P[N = n] = \sum_{x=0}^{x=n} P[X = x]\, P[Y = n-x / X = x]$ et en déduire une expression de cette probabilité.

3) On montre que N a pour moyenne $15p+5$ et pour variance $(1-p)(\frac{15}{4} + \frac{100}{11} p)$ et que N suit une loi normale.
Calculer la valeur de p telle que $P[N \geq 10] = 0{,}95$ et conclure.

Exercice 11

Une urne contient N boules numérotées de 1 à N. On tire au hasard, successivement et sans remise, deux boules de cette urne. On note X (resp. Y) la variable aléatoire égale au numéro porté par la première (resp. deuxième) boule tirée.
1)Déterminer la loi de X. Calculer $\mathbf{E}(X)$ et $\mathbf{V}(X)$.
2)Déterminer la loi du couple (X,Y).
3)Déterminer la loi de Y sachant que $X = k$ (k dans $\{1, \ldots, N\}$). En déduire la loi de Y.
4)On pose $U = \mathrm{Max}(X,Y)$, $V = \mathrm{Min}(X,Y)$.
a)Calculer $P[U < k]$, lorsque k est dans $\{2, \ldots, N+1\}$. En déduire la loi de U.
b)Déterminer la loi du couple (U,V). En déduire la loi de V.

c)Déterminer la loi de $Z = U - V$.

Exercice 12

Un contrôleur du service des impôts doit visiter pendant les 2 jours de week-end n domiciles choisis au hasard d'une localité. On admet qu'il y a une même probabilité p que le locataire du domicile soit présent lors de son passage et ce, quels que soient le domicile à visiter et le jour de son passage, et que l'absence ou la présence du locataire d'un domicile i ne dépend pas de l'absence ou de la présence du locataire d'un domicile j $(i \neq j)$.

a)Soit X la variable aléatoire égale au nombre de domiciles que le contrôleur peut visiter (les locataires de ces domiciles sont présents) lors de son premier passage.
Quelle est la loi de probabilités de X ?
Calculer à 0,001 près les probabilités $P[X = k]$ pour $k = 4$ et 5, $n = 6$ et $p = 0{,}8$.

b) Si, au premier passage, le contrôleur n'a pas pu visiter tous les n domiciles, il repasse une 2ème fois auprès des $(n - X)$ domiciles qu'il n'a pas pu visiter lors de son 1er passage.
Soit Y la VAR. nombre de domiciles visités lors du 2ème passage. Si $X = n$, on pose $Y = 0$. On définit $Z = X + Y$.
Calculer, pour $n = 6$ et $p = 0{,}8$ les probabilités suivantes :

i) $P[Y = 1 / X = 0]$ et $P[Y = 0 / X = 1]$;
ii) $P([X = 0] \cap [Y = 1])$ et $P([X = 1] \cap [Y = 0])$;
iii) $P[Z = 1]$, $P[Z = 1 / X = 1]$ et $P[X = 1 / Z = 1]$.

c)Montrer que la loi de probabilité de Y conditionnelle à $[X = x]$ pour x dans $\{0,1,\ldots,n-1\}$ est binomiale de paramètres $(n - x)$ et p .
En déduire la loi de probabilités conjointe de X et Y, est-ce que X et Y sont indépendantes ?
Calculer la covariance de X et Y pour $n = 3$ et $p = 0{,}8$.

d) Montrer que pour $0 \leq h \leq k \leq n$, $C_n^h C_{n-h}^{k-h} = C_n^k C_k^h$..

e)Pour k dans $\{0,1,\ldots,n\}$, exprimer l'événement $[Z = k]$ en fonction des événements $[X = x] \cap [Y = k - x]$ pour $0 \leq x \leq k$. Déduire de cette relation, de c) et de d) que Z est une VAR binomiale dont on donnera une interprétation des paramètres.