Vecteurs aléatoires

- distribution conjointe 2 variables aléatoires
- masse conjointe | densité conjointe
- distributions : marginales | conditionnelles
- indépendance en probabilité

Étude entre 2 variables

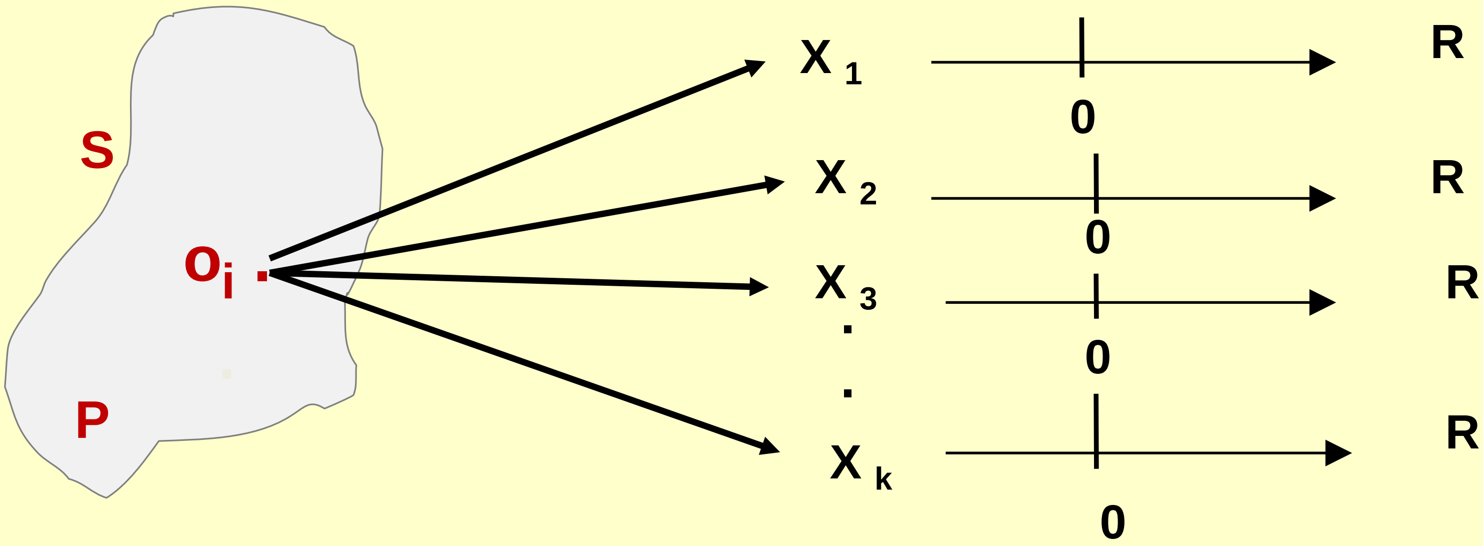
- 2 variables quantitatives : covariance et corrélation
- 2 variables catégoriques : coefficient contingence
- 1 variable continue + 1 var catégorique :

Étude d'un système multivariable

- linéaire | non linéaire
- formule propagation de la variabilité
- formules combinaisons linéaire
- modèle de l'échantillonnage

Vecteurs aléatoires

S : espace de probabilités



P : fonction de probabilité

Fonction de répartition conjointe

(k = 2)

$$F_{X1,\,X2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$$

fonction de masse probabilité

p : 2 variables discrètes

fonction de densité probabilité

f : 2 variables continues

Exemple 1 caractéristiques de voitures (1993)

27 variables 93 voitures

id	Manufac Turer X1	Model X2	Cate Gory X3	Price Y	City MPG X4	Highway MPG X5	Engine Size X6	Horse Power X7	Fuel Tank X8	Passen Gers X9	Weight X10	.
Acura_in	Acura	Integra	Small	15,9	25	31	1,8	140	13,2	5	2705	.
Acura_le	Acura	Legend	Midsize	33,9	18	25	3,2	200	18,0	5	3560	.
Audi_90	Audi	90	Compact	29,1	20	26	2,8	172	16,9	5	3375	.
Audi_100	Audi	100	Midsize	37,7	19	26	2,8	172	21,1	6	3405	.
BMW_535	BMW	535i	Midsize	30,0	22	30	3,5	208	21,1	4	3640	.
Buick_ce	Buick	Century	Midsize	15,7	22	31	2,2	110	16,4	6	2880	.
.	.	.	.	.	.	.	.	.	.	.	.	.
Volvo_850	Volvo	850	Midsize	26,7	20	28	2,4	168	19,3	5	3245	.

Comment visualiser toutes les données?

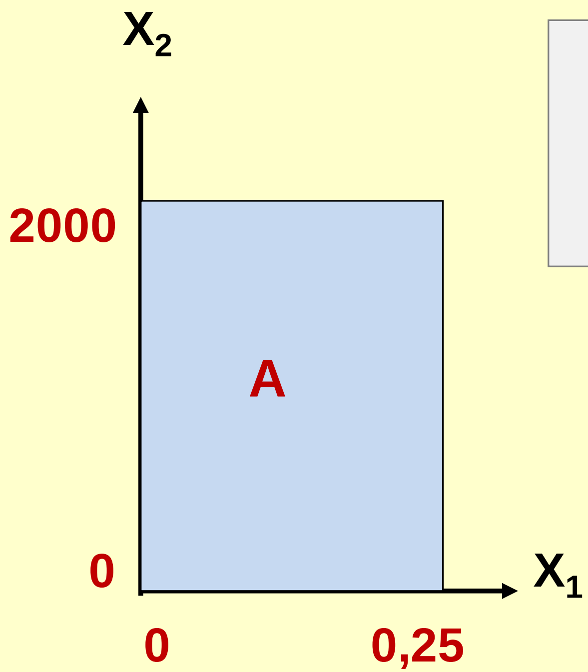
Le prix Y est-il relié aux X_i? lesquelles? comment?

probabilité conjointe de 2 variables aléatoires continues

Exemple 2 : **2 variables continues**

X_1 : diamètre de soudures $\quad 0 \leq x_1 \leq 0{,}25$ (po)

X_2 : résistance au cisaillement $\quad 0 \leq x_2 \leq 2000$ (lb)



modèle proposé : distribution uniforme

densité uniforme

$$f_{X1,\,X2}(x_1, x_2) = 1/500 \quad \text{si} \quad (x1, x2) \in A$$
$$= 0 \qquad \text{autrement}$$

densité conjointe : cas général

$$f_{X1,\,X2}(x_1, x_2) \geq 0$$

$$\iint f_{X1,\,X2}(x_1, x_2)\, dx_1\, dx_2 = 1$$

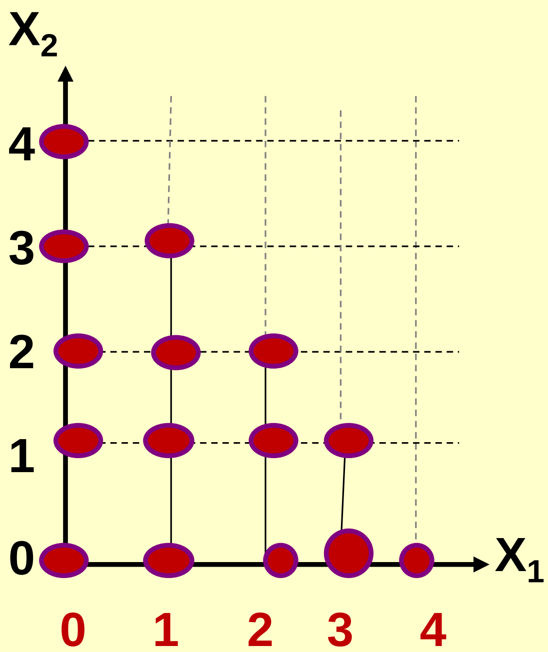
Exemple 3 : 2 variables discrètes – examen pièce

nombre de défauts majeurs / mineurs selon 2 critères

X_1 : nombre défauts mineurs $x_1 = 0, 1, 2, 3, 4$

X_2 : nombre défauts majeurs $x_2 = 0, 1, 2 ,3, 4$

avec la contrainte $0 \leq X_1 + X_2 \leq 4$



X1	X2	Prob
0	0	1/30
0	1	1/30
0	2	1/30
0	3	1/30
0	4	3/30
1	0	1/30
1	1	1/30
1	2	2/30
1	3	3/30

X1	X2	Prob
2	0	2/30
2	1	3/30
2	2	3/30
3	0	3/30
3	1	4/30
4	0	1/30

Exemple 4: assemblage de 2 composants

exemple traité dans chapitre **PROBABILITÉS**

X_1:composant 1 : possibilité de i = 0 -1- 2 -3 défauts

X_2:composant 2 : possibilité de j= 0 -1- 2 -3 - 4 défauts

modèle proposé : probabilité pour l'assemblage

(i , j) = dénouement 4*5 = 20 cas

$P(X_1=0, X_2=0) = 0.5$ assemblage sans défaut

$P(X_1=i, X_2=j) = k / (i + j)$ si (i, j) = (0, 1), ….. (3, 4)

$k = ?$

Exemple 4

P (X1=0, X2=0) = 0,5

Autres cas

si (i, j)

= (0, 1), ….. (3, 4)

$P\ (X_1=i, X_2=j\ )$

$= k\ /\ (\ i + j\ )$

k = ?

i	j	i + j	k / (i+j)	Prob (i, j)
0	0	0		0,500
0	1	1	k/1	0,072
0	2	2	k/2	0,036
0	3	3	k/3	0,024
0	4	4	k/4	0,018
1	0	1	k/1	0,072
1	1	2	k/2	0,036
1	2	3	k/3	0,024
1	3	4	k/4	0,018
1	4	5	k/5	0,014
2	0	2	k/2	0,036
2	1	3	k/3	0,024
2	2	4	k/4	0,018
2	3	5	k/5	0,014
2	4	6	k/6	0,012
3	0	3	k/3	0,024
3	1	4	k/4	0,018
3	2	5	k/5	0,014
3	3	6	k/6	0,012
3	4	7	k/7	0,010
		total	k * 6,91	1,000

= 0,5

= k*6,91

k = 0,072

Exemple 4 suite réorganisation en tableau croisé

probabilités conjointes

(X1, X2)	X2 = 0	1	2	3	4	total
X1 = 0	0,500	0,072	0,036	0,024	0,018	0,651
1	0,072	0,036	0,024	0,018	0,014	0,165
2	0,036	0,024	0,018	0,014	0,012	0,105
3	0,024	0,018	0,014	0,012	0,01	0,079
total	0,633	0,151	0,093	0,069	0,055	1,000

PROBABILITÉS MARGINALES - PROBABILITÉS CONDITIONNELLES

conjointe $\longrightarrow$ marginales

sommation $p(X_1) = \sum p(X_1,X_2)$

$p(X_2) = \sum p(X_1,X_2)$

conjointe $\longrightarrow$ conditionnelles

normalisation $p(X_1 \mid X_2) = p(X_1,X_2) / p(X_2)$

$p(X_2 \mid X_1) = p(X_1,X_2) / p(X_1)$

? marginales $\longrightarrow$ conjointe

oui **si** indépendance de X_1 et X_2

non **si** X1 et X2 sont dépendantes

Exemple 4: suite probabilités marginales

(X1,X2)	X2 = 0	X2 =1	X2 = 2	X2 = 3	X2 = 4	total
X1 = 0	0,500	0,072	0,036	0,024	0,018	0,651
X1 = 1	0,072	0,036	0,024	0,018	0,014	0,165
X1 = 2	0,036	0,024	0,018	0,014	0,012	0,105
X1 = 3	0,024	0,018	0,014	0,012	0,010	0,079
total	0,633	0,151	0,093	0,069	0,055	1,000

$P_{x2}(x2)$

$P_{x1}(x1)$

Exemple 4 : suite probabilités conditionnelles

X2/X1	X2 = 0	X2 =1	X2 = 2	X2 = 3	X2 = 4	tot
X1=0	0,77	0,11	0,06	0,04	0,03	1
X1=1	0,44	0,22	0,15	0,11	0,09	1
X1=2	0,34	0,22	0,17	0,14	0,11	1
X1=3	0,30	0,23	0,18	0,15	0,13	1

chaque ligne est une distribution conditionnelle

chaque colonne est une distribution conditionnelle

X1/X2	X2 =0	X2 =1	X2 =2	X2 =3	X2 =4
X1=0	0,79	0,48	0,39	0,35	0,33
X1=1	0,11	0,24	0,26	0,23	0,26
X1=2	0,06	0,16	0,14	0,21	0,22
X1=3	0,04	0,12	0,16	0,17	0,19
total	1	1	1	1	1

INDÉPENDANCE (en probabilité) de 2 variables

X_1 et X_2 sont indépendantes si

$p_{X1,X2}(x_1,x_2) = p_{X1}(x_1)\ p_{X2}(x_2)$ **variables discrètes**

$f_{X1,X2}(x_1,x_2) = f_{X1}(x_1)\ f_{X2}(x_2)$ **variables continues**

loi conjointe = produit des lois marginales

Conséquence

toutes les lois conditionnelles sont identiques

Exemple 4: assemblage de 2 composants - indépendance ?

(X_1,X_2)	X2=0	X2=1	X2=2	X2=3	X2=4	$p(X_1)$
X1=0	0,500	0,072	0,036	0,024	0,018	0,651
X1=1	0,072	0,036	0,024	0,018	0,014	0,165
X1=2	0,036	0,024	0,018	0,014	0,012	0,105
X1=3	0,024	0,018	0,014	0,012	0,010	0,079
$p(X_2)$	0,633	0,151	0,093	0,069	0,055	1,000

p(0, 0) = 0,500

≠

p(X1=0) * p(X2=0) = 0,651 * 0,633 = 0,412

X1 et X2 pas indépendants = X1 et X2 dépendants

comment mesurer degré dépendance? …. plus loin

autre cas

p(2,3) = 0,014 ≠ 0,0072 = 0,105 * 0,069 = p(X1=2) * p(X2=3)

Exemple 4: suite

modèle avec indépendance de X1 et X2

(X_1,X_2)	X2=0	X2=1	X2=2	X2=3	X2=4	$p_{X1}(X_1)$
X1=0	0,412	0.098	0,061	0,045	0,036	0,651
X1=1	0,105	0,025	0,015	0,011	0,099	0,165
X1=2	0,066	0,016	0,010	0,007	0,006	0,105
X1=3	0,050	0,012	0,007	0,005	0,004	0,079
$p_{X2}(X_2)$	0,633	0,151	0,093	0,069	0,055	1,000

modèle postulé = dépendance

(X_1,X_2)	X2=0	X2=1	X2=2	X2=3	X2=4	$p_{X1}(X_1)$
X1=0	0,500	0,072	0,036	0,024	0,018	0,651
X1=1	0,072	0,036	0,024	0,018	0,014	0,165
X1=2	0,036	0,024	0,018	0,014	0,012	0,105
X1=3	0,024	0,018	0,014	0,012	0,010	0,079
$p_{X2}(X_2)$	0,633	0,151	0,093	0,069	0,055	1,000

exemple 5 : modélisation avec indépendance prototype véhicule fin de test (essai)

M1 : moteur OK M2 : moteur état moyen M3 : moteur état médiocre

Probabilité = 0,6 = 0,3 = 0,1

D1 : direction OK D2 : direction non opérationnelle

Probabilité : = 0,.8 = 0,2

A1 : amortisseur OK A2 : amortisseur dégradé A3 : amortisseur brisé

Probabilité = 0,7 = 0,2 = 0,1

probabilités conjointes avec hypothèse d'indépendance

Mi ∩	Dj ∩	Ak	prob	Mi ∩	Dj ∩	Ak	prob	Mi ∩	Dj ∩	Ak	prob
1	1	1	0,336	2	1	1	0,168	3	1	1	0,056
1	1	2	0,096	2	1	2	0,048	3	1	2	0,016
1	1	3	0,048	2	1	3	0,024	3	1	3	0,008
1	2	1	0,084	2	2	1	0,042	3	2	1	0,014
1	2	2	0,024	2	2	2	0,012	3	2	2	0,004
1	2	3	0,012	2	2	3	0,006	3	3	3	0,002

Exemple 5 (suite) : prototype véhicule fin de test (essai)

probabilités conjointes avec hypothèse d'indépendance

Mi ∩	Dj ∩	Ak	prob	Mi ∩	Dj ∩	Ak	prob	M i ∩	Dj ∩	Ak	prob
1	1	1	0,336	2	1	1	0,168	3	1	1	0,056
1	1	2	0,096	2	1	2	0,048	3	1	2	0,016
1	1	3	0,048	2	1	3	0,024	3	1	3	0,008
1	2	1	0,084	2	2	1	0,042	3	2	1	0,014
1	2	2	0,024	2	2	2	0,012	3	2	2	0,004
1	2	3	0,012	2	2	3	0,006	3	3	3	0,002

F_k : k composants sont fonctionnels k = 0, 1, 2, 3 P(Fk) = ?

$P(F_3) = P(M1 \cap D1 \cap A1) = 0{,}336$

$P(F_2) = P(M1 \cap D\,1 \cap A2) + P(113) + P(211) + P(311) = 0{,}452$

$P(F_1) = P(122) + P(123) + P(212) + P(213) + P(221) + P(312) + P(313) + P(321) = 0{,}188$

$P(F_0) = P(222) + P(223) + P(322) + P(323) = 0{,}024$

MESURES pour décrire le degré de liaison

entre 2 variables quantitatives (continues ou discrètes)

COVARIANCE ET CORRÉLATION

Covariance entre les variables X_1 et X_2 quantitatives

$$\text{Cov}(X_1, X_2) = \sigma_{12}$$

$$= E[(X_1 - E(X_1)) * (X_2 - E(X_2))]$$

$$= E(X_1 * X_2) - E(X_1)E(X_2)$$ **pour le calcul**

Coefficient de corrélation entre les variables X_1 et X_2

$$\text{Corr}(X_1,X_2) = \rho_{12} = \sigma_{12} / \sigma_1 \sigma_2$$

σ_{12} = covariance de (X_1,X_2) = Cov (X_1, X_2)

σ_1 = écart type de X_1 = ET(X_1)

σ_2 = écart type de X_2 = ET(X_2)

Propriétés du coefficient de corrélation linéaire

1. $-1 \leq \rho_{12} \leq 1$ nombre sans unité de mesure

2. Si X_1 et X_2 sont indépendantes alors $\rho_{12} = 0$

3. La proposition inverse de (2) est fausse :

   possible que $\rho_{12} = 0$ sans que les variables soient indépendantes

4. Si $X_2 = a + b X_1$ alors $\rho_{12} = 1$ si $b > 0$ et $\rho_{12} = -1$ si $b < 0$

   Si $\rho_{12} = \pm 1$ alors $X_2 = a + b X_1$

   ρ_{12} mesure de dépendance fonctionnelle LINÉAIRE entre 2 variables

5. $100^*\rho_{12}^2$ = % de variation de X_2 expliqué par X_1 (ou X_1 par X_2)

ρ_{12}	-1	-0,9	-0,7	-0,5	-0,3	-0,1	0
ρ_{12}	1	0.9	0,7	0,5	0,3	0,1	0
%	100	81	49	25	9	1	0
Classification		forte	élevée	moyenne	faible	très faible	nulle

Propriétés du coefficient de corrélation linéaire

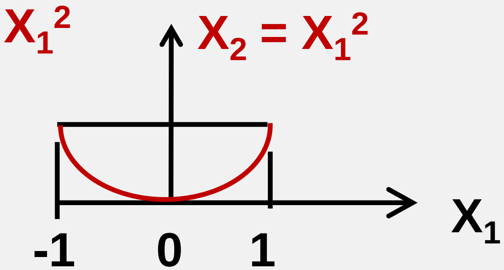
6. Lien entre coefficient de corrélation et fonction mathématique ?

En général, la valeur de ρ_{X1X2} apporte AUCUNE information sur la fonction g pouvant liée X_1 et X_2 $\quad X_2 = g(X_1)$

exception: si $\rho_{X1X2} = \pm 1$ alors $X_2 = a + b\, X_1$ (énoncé (4) p.18)

Exemple: soit $X_2 = X_1^2$ et X_1 variable distribution symétrique par rapport à 0: alors $\rho_{X1X2} = 0$

a) X_1 loi uniforme sur l'intervalle (-1, 1) $\quad X_2 = X_1^2$



$E(X_1) = 0$ et $E(X_1^3) = 0$

$\mathrm{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$

$= E(X_1X_2) = E(X_1^3) = 0$

$\rho_{X1X2} = 0$

b) X_1 loi uniforme sur l'intervalle (0, 1) $\quad X_2 = X_1^2$

$E(X_1) = 1/2 \quad E(X_1^2) = 1/3 \quad E(X_1^3) = ¼ \quad E(X_1^4) = 1/5$

$\sigma_1 = \mathrm{sqrt}[1/3 - (1/2)(1/2)] = \mathrm{sqrt}(1/12)$

$X_2 = X_1^2 \quad E(X_2) = 1/3 \quad \sigma_2 = \mathrm{sqrt}\,[1/5 - (1/3)(1/3)] = \mathrm{sqrt}(4/45)$

$\mathrm{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = E(X_1^3) - (1/2)(1/3) = 1/12$

$\rho_{X1X2} = (1/12) / \mathrm{sqrt}(1/12)\mathrm{sqrt}(4/45) = 0{,}0833 / 0{,}0{,}861 = 0{,}967$

COVARIANCE ET CORRÉLATION

$E(X_1)$ = moyenne de $X_1 = \mu_{x1}$ $E(X_2)$ = moyenne de $X_2 = \mu_{x2}$

$ET(X_1)$ = écart type de $X_1 = \sigma_{x1}$ $ET(X_2)$ = écart type de $X_2 = \sigma_{x2}$

$$\text{Covariance}\ (X_1, X_2) = E\,[(X_1 - \mu_{x1})\ (X_2 - \mu_{x2})\ ] = \sigma_{X1\ X2}$$

$$\text{Corrélation}\ (X_1, X_2) = \rho_{X1X2} = \sigma_{X1X2} / \sigma_{x1}\ \sigma_{x2}$$

$$= E[(X_1 - \mu_{x1})/\sigma_{x1})\ (X_2 - \mu_{x2})/\sigma_{x2})]$$

Posons $Z_1 = (X_1 - \mu_{x1}) / \sigma_{x1}$ variable centrée réduite de X_1

$Z_2 = (X_2 - \mu_{x2}) / \sigma_{x2}$ variable centrée réduite de X_2

Alors $E(Z_1) = \mu_{Z1} = 0$ $E(Z_2) = \mu_{Z2} = 0$

$ET(Z_1) = \sigma_{Z1} = 1$ $ET(Z_2) = \sigma_{Z2} = 1$ $\sigma_{Z1} * \sigma_{Z2} = 1$

$$\text{Covariance}\ (Z_1, Z_2) = \sigma_{Z1Z2} = E[(Z_1 - \mu_{Z1})(Z_2 - \mu_{Z2})] = E(Z_1 Z_2)$$

$$\text{Corrélation}\ (Z_1, Z_2) = \rho_{Z1Z2} = \sigma_{Z1Z2} / \sigma_{Z1}\ \sigma_{Z2} = \sigma_{Z1Z2}$$

$$= \text{Covariance}(Z_1, Z_2)$$

Corrélation (X_1, X_2) = Corrélation (Z_1, Z_2)

Exemple 4: suite assemblage de 2 composants

calcul de la covariance et de la corrélation

X1 et X2 variables ne sont pas indépendantes (p.13)

(X1,X2)	X2=0	X2=1	X2=2	X2=3	X2=4	$p(X_1)$
X1=0	0,500	0,072	0,036	0,024	0,018	0,651
X1=1	0,072	0,036	0,024	0,018	0,014	0,165
X1=2	0,036	0,024	0,018	0,014	0,012	0,105
X1=3	0,024	0,018	0,014	0,012	0,010	0,079
$p(X_2)$	0,633	0,151	0,093	0,069	0,055	1,000

$E(X_1)$ = 0,612 $E(X_2)$ = 0,764 $E(X1X2)$ = 0,874

$ET(X_1)$ = 0,959 $ET(X_2)$ = 1,20

$Cov (X_1,X_2)$ = 0,874 – 0,612*0,74 = 0,406

$Corr(X_1, X_2)$ = 0,406 / (0,959*1,20) = 0,35

corrélation force moyenne

ÉTUDE de 2 variables catégoriques

X et Y variables catégoriques

modalités de X : $A_1, A_2, \ldots, A_r$ modalités de Y : $B_1, B_2, \ldots, B_c$

effectifs conjoints (n_{ij})

X \ Y	B_1	B_2	...	B_j	...	B_c	TOTAL
A_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	n_{1+}
A_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	n_{2+}
.	.	.		.		.	.
A_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	n_{i+}
.	.	.		.		.	.
A_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	n_{r+}
TOTAL	n_{+1}	n_{+2}	...	n_{+j}	...	n_{+c}	n

n_{ij} = EFFECTIF CONJOINT de (A_i, B_j)

$n_{i+} = \sum_j n_{ij}$ = effectif de A_i,

$n_{+j} = \sum_i n_{ij}$ = effectif de B_j,

$f_{ij} = n_{ij}/n$ = fréquence conjointe de (A_i, B_j)

$f_{i+} = n_{i+}/n$ = fréquence marginale de A_i

$f_{+j} = n_{+j}/n$ = fréquence marginale de B_j

$\frac{f_{ij}}{f_{+j}} = \frac{n_{ij}}{n_{+j}}$ = fréquence conditionnelle de A_i si $Y = B_j$

$\frac{f_{ij}}{f_{i+}} = \frac{n_{ij}}{n_{i+}}$ = fréquence conditionnelle de B_j si $X = A_i$

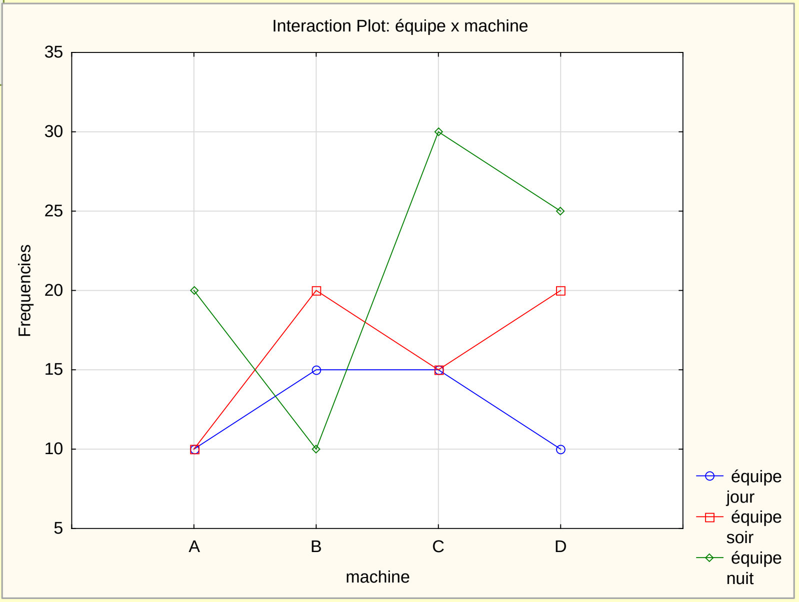
f_{ij} peut servir pour définir un modèle de probabilités p_{ij}

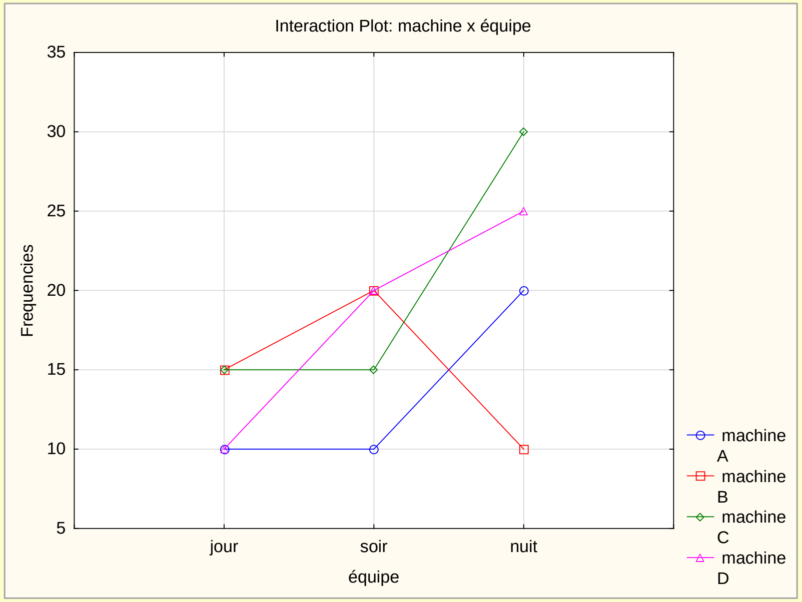
ÉTUDE de 2 variables catégoriques

Exemple: nombre d'arrêts observés pendant 365 jours
4 machines (A, B, C, D) X 3 équipes (jour, soir, nuit)

machine	équipe jour	équipe soir	équipe nuit	Row Totals
A	10	10	20	40
B	15	20	10	45
C	15	15	30	60
D	10	20	25	55
All Grps	50	65	85	200

	machine	équipe jour	équipe soir	équipe nuit	Row Totals
Count	A	10	10	20	40
Column Percent		20,0%	15,4%	23,5%	
Row Percent		25,0%	25,0%	50,0%	
Total Percent		5,0%	5,0%	10,0%	20,0%
Count	B	15	20	10	45
Column Percent		30,0%	30,8%	11,8%	
Row Percent		33,3%	44,4%	22,2%	
Total Percent		7,5%	10,0%	5,0%	22,5%
Count	C	15	15	30	60
Column Percent		30,0%	23,1%	35,3%	
Row Percent		25,0%	25,0%	50,0%	
Total Percent		7,5%	7,5%	15,0%	30,0%
Count	D	10	20	25	55
Column Percent		20,0%	30,8%	29,4%	
Row Percent		18,2%	36,4%	45,5%	
Total Percent		5,00%	10,0%	12,5%	27,5%
Count	All Grps	50	65	85	200
Total Percent		25,0%	32,5%	43%	





ÉTUDE de 2 variables catégoriques

Comment mesurer de degré de liaison entre les 2 variables?

$$D^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad \text{où } e_{ij} = \frac{n_{i+} * n_{+j}}{n}$$

coefficient de contingence

mesure de degré de dépendance entre X et Y

$C = [\, D^2 / (D^2 + n) \,]^{0,5}$ varie entre 0 et 1

rôle analogue à r

exemple

statistique	valeur
Pearson Chi-square	12,02
Contingency coefficient	0,238

ÉTUDE de 2 variables
X: variable catégorique Y : variable continue

étude dans le chapitre de l'analyse de la variance

exemple

	22 X_type revêtement	23 Y_conductivité
1	A	143
2	A	141
3	A	150
4	A	146
5	B	152
6	B	149
7	B	137
8	B	143
9	C	134
10	C	133
11	C	132
12	C	127
13	D	129
14	D	127
15	D	132
16	D	129
17	E	147
18	E	148
19	E	144
20	E	142

PROCESSUS / SYSTÈME

FACTEURS $X_1, X_2, \ldots, X_K$: variables contrôlées
que l'on fait varier

X_1
X_2
. . .
X_k

SYSTÈME

réponse Y
sortie mesurée

facteurs de bruit
$Z_1, Z_2, \ldots$
connus mais non contrôlés

ε
erreur expérimentale:
variablilité source inconnue

<u>étude performance processus : à faire avec des données !</u>

mode actif : expérimentation - plan de collecte à concevoir

mode passif : observations historiques / observationnelles

MODÉLISATION STATISTIQUE

Toute analyse statistique repose sur un modèle :

- fonction f pour représenter une relation entre input X et output Y
- hypothèse distributionnelle pour le terme d'erreur ε

X_1

X_2

…

X_k

$\longrightarrow$ **PROCESSUS** $\longrightarrow$ $Y = f(X)$

ε $\quad$ $\varepsilon \sim N(0,\sigma^2)$

$$Y = f(X_1, X_2, \ldots, X_k ; \beta_0, \beta_1, \beta_2, \ldots) + \varepsilon$$

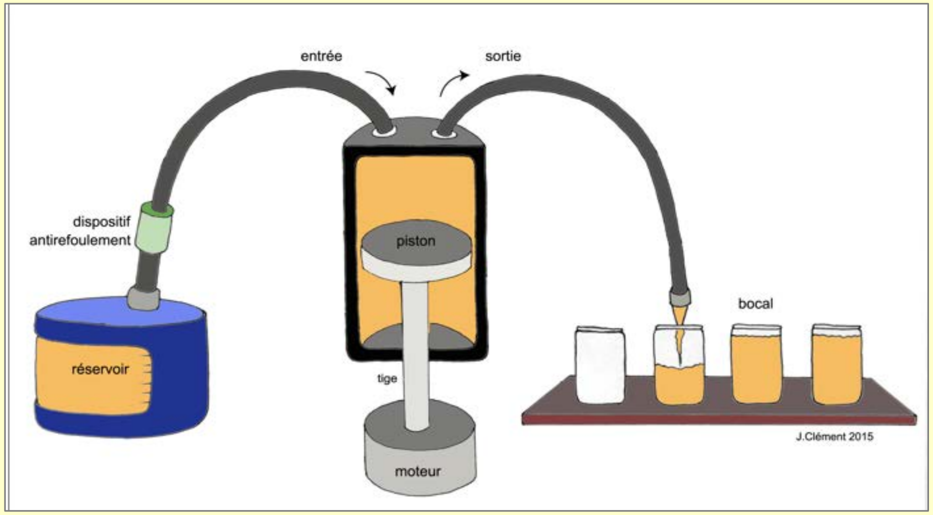
f : fonction connue - cas assez rare - simulation

f : inconnue $\longrightarrow$ approximation polynôme

$\beta_0, \beta_1, \beta_2, \ldots$: paramètres statistiques inconnus

système avec variables aléatoires

Exemple



$$D = (0{,}0157 * R * R * L - B) * (V/60)$$

D : débit du fluide	(ml / sec)
R : rayon du piston	(mm)
L = longueur du bras	(mm)
V : vitesse du moteur	(rpm)
B = refoulement (« backflow »)	(ml)

système

Soit $W = f(X_1, X_2, \ldots, X_n)$ $\quad X_1, X_2, \ldots, X_n$ variables aléatoires

avec $E(X_i) = \mu_i \quad Var(X_i) = \sigma_i^2 \quad cov(X_i, X_{i'}) = \rho_{ii'}\, \sigma_i \sigma_{i'}$

Alors $E(W) \approx f(\mu_1, \mu_2, \ldots \mu_n)$

$$Var(W) \approx \sum [\partial f / \partial X_i]^2 \sigma_i^2 + \sum \sum [\partial^2 f / \partial X_i \partial X_{i'}]^2 \rho_{ii'}\, \sigma_i \sigma_{i'}$$

Les dérivées sont évaluées à $(\mu_1, \mu_2, \ldots \mu_n)$.

Cas : combinaison linéaire

$b, a_1, a_2, \ldots, a_n$ constantes

cas 1 : $X_1, X_2, \ldots, X_n$ variables aléatoires indépendantes

$W = b + \sum a_i X_i$ combinaison linéaire

Théorème $E(X_i) = \mu_i$ moyenne de X_i

$VAR(X_i) = \sigma_i^2$ variance de X i

alors

$$E\left(\sum a_i X_i\right) = b + \sum a_i \mu_i$$

$$VAR\left(\sum a_i X_i\right) = \sum a_i^2 \sigma_i^2$$

Cas 2 : dépendance des X i ρ_{ij} = corrélation entre X i et X j

$$E\left(\sum a_i X_i\right) = b + \sum a_i \mu_i$$

$$VAR\left(\sum a_i X_i\right) = \sum a_i^2 \sigma_i^2 + \sum_{i \neq j}\sum a_i a_j \sigma_i \sigma_j \rho_{ij}$$

Exemple: combinaison linéaire

Exemple 5

p(X=6, Y = 4)

= p(X=6)*p(Y=4)

= 0,20* 0,15

= 0,0300

		probabilités conjointes p(x,y) modèle : X, Y variables indépendantes							
	X	3	5	6	9	10	13	14	
Y	prob	0,05	0,15	0,20	0,30	0,10	0,10	0,10	1,00
1	0,05	0,0025	0,0075	0,0100	0,0150	0,0050	0,005	0,005	0,05
2	0,10	0,0050	0,0150	0,0200	0,0300	0,0100	0,010	0,010	0,10
3	0,30	0,0105	0,0450	0,0600	0,0900	0,0300	0,030	0,030	0,30
4	0,15	0,0075	0,0225	0,0300	0,0450	0,0150	0,015	0,015	0,15
5	0,20	0,0100	0,0300	0,0400	0,0600	0,0200	0,020	0,020	0,20
6	0,20	0,0100	0,0300	0,0400	0,0600	0,0200	0,020	0,020	0,20
	1,00	0,05	0,15	0,20	0,30	0,10	0,10	0,10	1,00

moyenne et variance de W = 4X + 3Y ?

Exemple 5

			valeurs de W = 4 X + 3 Y						
		X	3	5	6	9	10	13	14
Y	3Y	4X	12	20	24	36	40	52	56
1	3		15	23	27	39	43	55	59
2	6		18	26	30	42	46	58	62
3	9		21	29	33	45	49	61	65
4	12		24	32	36	48	52	64	68
5	15		27	35	39	51	55	67	71
6	18		30	38	42	54	58	70	74

Liste des valeurs ordonnées de W

15 18 21 23 24 26 **27** 29 **30** 32 33 35 36 38 **39**
42 43 45 46 48 49 51 52 54 **55** **58** 59 61 62 64
65 67 68 70 71 74

valeurs noires: plusieurs fois dans le tableau

		CALCUL DES MOYENNES							
	a=4	x=3	5	6	9	10	13	14	E(X
b=3	45,85	0,15	0,75	1,20	2,70	1,00	1,30	1,40	= 8,50
y=1	0,05	0,0375	0,173	0,270	0,585	0,215	0,275	0,295	
2	0,20	0,090	0,390	0,600	1,260	0,460	0,580	0,620	
3	0,90	0,315	1,305	1,980	4,050	1,470	1,830	1,950	
4	0,60	0,180	0,720	1,080	2,160	0,780	0,960	1,020	
5	1,00	0,270	1,050	1,560	3,060	1,100	1,340	1,420	
6	1,20	0,300	1,140	1,680	3,240	1,160	1,400	1,480	
E(Y)	= 3,95							E(W)	= 45,85

$a_{i\,j}$

$a_{i\,j} = (4x_i + 3y_j)^* \, p(x_i , y_j)$ → Calcul direct : $E(W) = \sum \sum a_{i\,j}$

Calcul formule : E(W) = 4*E(X) + 3*E(Y) = 4*8,50 + 3*3,95 = 45,85

		CALCUL DES VARIANCES							
	a=4	3	5	6	9	10	13	14	V(X)
b=3	178,528	1,513	1,838	1,250	0,075	0,225	2,025	3,025	=9,95
1	0,435	2,379	3,916	3,553	0,704	0,041	0,419	0,865	
2	0,380	3,878	5,910	5,024	0,445	0,000	1,476	2,608	
3	0,271	9,263	12,777	9,907	0,065	0,298	6,886	11,002	
4	0,000	3,581	4,316	2,911	0,208	0,567	4,941	7,359	
5	0,221	3,553	3,532	1,877	1,591	1,674	8,946	12,650	
6	0,841	2,512	1,849	0,593	3,985	2,952	11,664	15,848	
V(Y)	= 2,148							V(W) =	178,528

b_{ij}

$b_{ij} = [w_{ij} - E(W)]^2 \, p(x_i, y_j)$ $\longrightarrow$ Calcul direct : $Var(W) = \sum \sum b_{ij}$

Calcul avec la formule : $Var(W) = 4^2 * 9{,}95 + 3^2 * 2{,}148 = 178{,}528$

Combinaison linéaire de variables aléatoires indépendantes

Application importante : processus d'échantillonnage

Définition d'un échantillon aléatoire: n variables X_i

(a) $X_1, X_2, X_3, \ldots\ldots\ldots, X_n$ indépendantes

(b) toutes les variables ont la même distribution que X

La **variable aléatoire moyenne**, notée $\overline{X}$,

est une combinaison linéaire des X_i

$$\overline{X} = \sum_{1}^{n} \frac{1}{n} X_i = \sum a_i X_i \qquad a_i = 1/n$$

Propriétés de $\overline{X}$

$$E[\overline{X}] = \frac{1}{n} \sum_{1}^{n} E[X] = E[X] = \mu$$

$$Var[\overline{X}] = \left(\frac{1}{n}\right)^2 \sum_{1}^{n} Var[X] = \frac{Var[X]}{n} = \frac{\sigma^2}{n}$$