IV) Variables aléatoires discrètes

Exemple : Une urne contient 1 boule blanche, 3 boules rouges et 6 boules noires.
On gagne 100 quand on tire une boule blanche
20 quand on tire une boule rouge et
0 quand on tire une boule noire.
Le gain est une variable aléatoire.

Définition : Soit Ω un ensemble d'évènements élémentaires sur lequel on a défini une probabilité. Faisons correspondre un nombre à chaque évènement élémentaire.
On définit ainsi une variable aléatoire discrète X.

Exemple :
X=Gain : {rouge, blanche, noire} → I**R**
rouge → 20
blanche → 100
noire → 0

Loi de probabilité d'une variable discrète

Exemple : X, la variable aléatoire gain, peut prendre les valeurs 0, 20 et 100 avec les probabilités suivantes :

P(X=0) = P(noire) = 0,6
P(X=20) = 0,3
P(X=100) = 0,1

Définition
L'association des valeurs x de la variable X et des probabilités P(X=x) est la loi de probabilité de X.

Exemple : loi de probabilité du Gain X

x	0	20	100
P(X=x)	0,6	0,3	0,1

Espérance mathématique :

$$E(X) = \sum_x xP(X = x)$$

Exemple : calculer le gain moyen. On l'appellera E(X) :

x	0	20	100
P(X=x)	0,6	0,3	0,1

$E(X) = 0\times 0{,}6 + 20\times 0{,}3 + 100\times 0{,}1 = 16$

Propriétés :

Soient a et b des constantes réelles, X et Y deux variables aléatoires :

- E(a X)=a E(X)
- E(b) = b
- E(X+Y) = E(X) + E(Y)

Exemples : (rappel E(X) = 16)

- E(2X+3) = E(2X) + E(3) = 2E(X) + 3 = 2x16 + 3 = 35

- Si Y telle que E(Y) = 8 alors E(X+Y) = E(X) +E(Y) = 16 + 8 = 24

Variance et écart type :

Définition

$V(X) = \sum_x (x - E(X))^2 P(X = x)$

L'écart type σ est la racine carrée de la variance

$$\sigma(X) = \sqrt{V(X)}$$

Propriété (à utiliser pour les calculs !)

$V(X) = E(X^2) - E(X)^2$ ou $E(X^2) = \sum_x x^2 P(X = x)$

Exemple : Calculer V(X) dans le cas de la variable aléatoire gain.

- On calcule d'abord le deuxième moment
$$deuxieme\ moment = E(X^2) = 0^2 \times 0{,}6 + 20^2 \times 0{,}3 + 100^2 \times 0{,}1$$
$$E(X^2) = 1120$$
- On en déduit la variance
$$V(X) = E(X^2) - (E(X))^2$$
$$V(X) = 1120 - 16^2 = 1120 - 256 = 864$$

Propriétés :

- V(a X) = a² V(X)
- V(X + b) = V(X)
- V(b) = 0   la variance d'une variable constante est toujours nulle
- V(X) ≥ 0
- V(X+Y) = V(X) + V(Y) dans le cas où les variables X et Y sont indépendantes

(Ce n'est pas toujours le cas !)

Exemples :

- V(3X+2) = V(3X) = $3^2 V(X)$= 9V(X) = 9 x 864 = 7776
- σ(2X+3) = $\sqrt{V(2X+3)} = \sqrt{V(2X)} = \sqrt{4V(X)} = 2\sqrt{V(X)} = 2\sigma(X) = 2\sqrt{864} \approx 58{,}8$

- Si Y est telle que E(Y)=20 et σ(Y)=10 et X et Y sont indépendantes,

Calculer σ(X+Y)
$$\sigma(X+Y) = \sqrt{V(X+Y)} = \sqrt{V(X)+V(Y)} = \sqrt{V(X)+\sigma(Y)^2} = \sqrt{864+100} = \sqrt{964}$$

Exercices sur les variables aléatoires :

Ex1 : a) Complétez les tableaux des lois de probabilité des variables suivantes :

x	-1	0	1
P(X=x)	0,2	0,5	0,3

P(X=1) = 1- (0,2 + 0,5) = 0,3

x	-1	0	1
P(Y = x)	0,25	0,5	0,25

P(Y = -1) = 1- (0,5 + 0,25)= 0,25

b) Calculez E(X) et E(Y)

$$E(X) = -1\times 0{,}2 + 0\times 0{,}5 + 1\times 0{,}3 = 0{,}1$$
$$E(Y) = -1\times 0{,}25 + 0\times 0{,}5 + 1\times 0{,}25 = 0$$

c) Calculez V(X) et V(Y)

$$E(X^2) = (-1)^2\times 0{,}2 + 0^2\times 0{,}5 + 1^2\times 0{,}3 = 0{,}5$$

d'où $V(X) = E(X^2) - E(X)^2 = 0{,}5 - 0{,}1^2 = 0{,}49$

$$E(Y^2) = (-1)^2\times 0{,}25 + 0^2\times 0{,}5 + 1^2\times 0{,}25 = 0{,}5$$

d'où $V(Y) = E(Y^2) - E(Y)^2 = 0{,}5 - 0^2 = 0{,}5$

Ex2 : Dans un magasin, on organise un jeu : chaque client lance un dé à 6 faces une seule fois et obtient alors une remise. S'il obtient 1, il obtient 10% de remise sur ses achats, s'il obtient 2,3,4 ou 5, il obtient 5% de remise, et s'il obtient 6 il obtient 20%.

a) Quelle est la loi de probabilité de la remise accordée ?
Soit X la remise accordée en %.

x	5	10	20
P(X=x)	$\frac{4}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

b) Quelle est la remise moyenne accordée au client ?

$$E(X) = 5\times\frac{4}{6} + 10\times\frac{1}{6} + 20\times\frac{1}{6} = \frac{50}{6} = \frac{25}{3} \cong 8{,}33$$

c) Proposer un jeu similaire tel que la remise pour un 6 soit de 50% mais que la remise moyenne soit de 10%

Notons r_i les remises des faces i (i= 1,2,3,4 ou 5)
On obtient alors la loi de X suivante :

x	r_1	r_2	r_3	r_4	r_5	50
P(X=x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Ecrivons l'équation donnée par $E(X) = 10$

$$E(X) = 10 = r_1\times\frac{1}{6} + r_2\times\frac{1}{6} + r_3\times\frac{1}{6} + r_4\times\frac{1}{6} + r_5\times\frac{1}{6} + 50\times\frac{1}{6}$$

$$\frac{1}{6}(r_1 + r_2 + r_3 + r_4 + r_5) = 10 - \frac{50}{6}$$

$$r_1 + r_2 + r_3 + r_4 + r_5 = \left(\frac{60}{6} - \frac{50}{6}\right)\times 6$$

$$r_1 + r_2 + r_3 + r_4 + r_5 = 10$$

On peut donc proposer des jeux qui répondent à la question :

- $r_1 = r_2 = r_3 = r_4 = r_5 = 2$
  la loi de X est alors

x	2	50
P(X=x)	$\frac{5}{6}$	$\frac{1}{6}$

- $r_1 = 10,\ r_2 = r_3 = r_4 = r_5 = 0$
  la loi de X est alors

x	0	10	50
P(X=x)	$\frac{4}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- … on peut en construire une infinité

Ex3 : Quelles sont les variables aléatoires de variance nulle ?
Cours : ce sont les variables constantes.

Ex 4 : On lance deux dés à 6 faces, on note leur somme.
On attribue un gain suivant la valeur obtenue avec la formule suivante
Gain= 2 fois la somme -10.
a) Quelle est la loi de probabilité du gain ?
b) Quel est le gain moyen ?
c) Proposer un autre gain donnant un jeu équilibré.

a) On se rappelle bien de l'exercice de probabilités… on a déjà les résultats sur la somme des deux dés :

	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

Somme	2	3	4	5	6	7	8	9	10	11	12
Proba	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

D'où le tableau suivant pour la variable gain :

Gain x= Somme fois 2 -10	-6	-4	-2	0	2	4	6	8	10	12	14
Proba	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

b) On peut alors calculer l'espérance du gain :

$$E(G) = \frac{-6-8-6+0+10+24+30+32+30+24+14}{36} = \frac{144}{36} = 4$$

Somme	2	3	4	5	6	7	8	9	10	11	12
Proba	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

c) Pour obtenir un jeu équilibré il faut que l'espérance soit nulle.

C'est le cas si :

par exemple en associant 0 à la somme 7

100 à la somme 2 et -100 à la somme 12 $\left(100\times\frac{1}{36}-100\times\frac{1}{36}=0\right)$

25 à la somme 3 et -25 à la somme 11 $\left(25\times\frac{2}{36}-25\times\frac{2}{36}=0\right)$

250 à la somme 4 et -250 à la somme 10

254321 à la somme 5 et -254321 à la somme 9

65725 à la somme 6 et -65725 à la somme 8

… à compléter

en utilisant la symétrie de la loi de la somme.

Plus simple : Soit S la variable de la somme des deux dés,

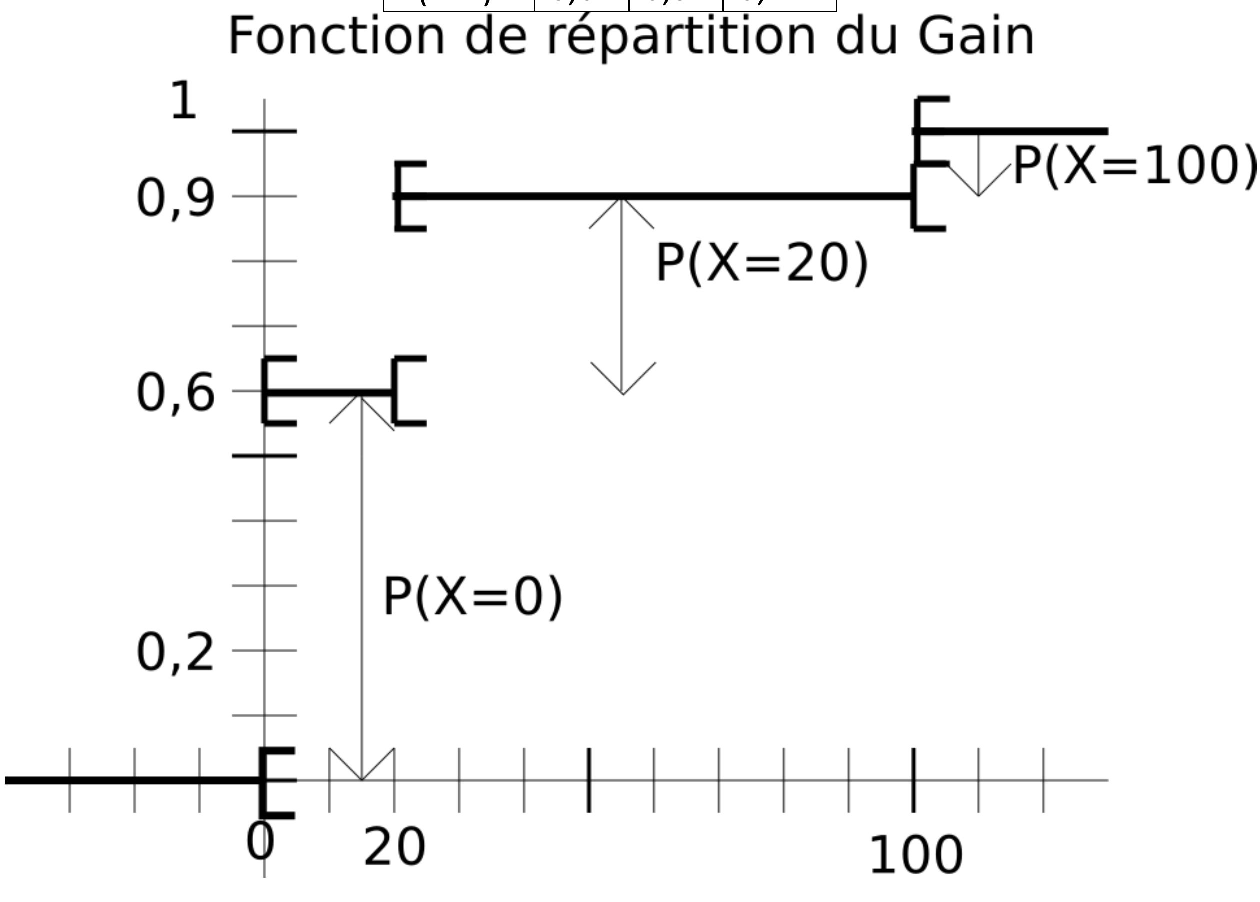
En posant G2 = S - E(S) on a E(G2)= E(S-E(S))=E(S)-E(E(S))=E(S)-E(S)=0

Ex 5 : Introduction à la notion de fonction de répartition

Exemple Cours: On va dessiner la fonction de répartition $F_X(t) = P(X \leq t)$

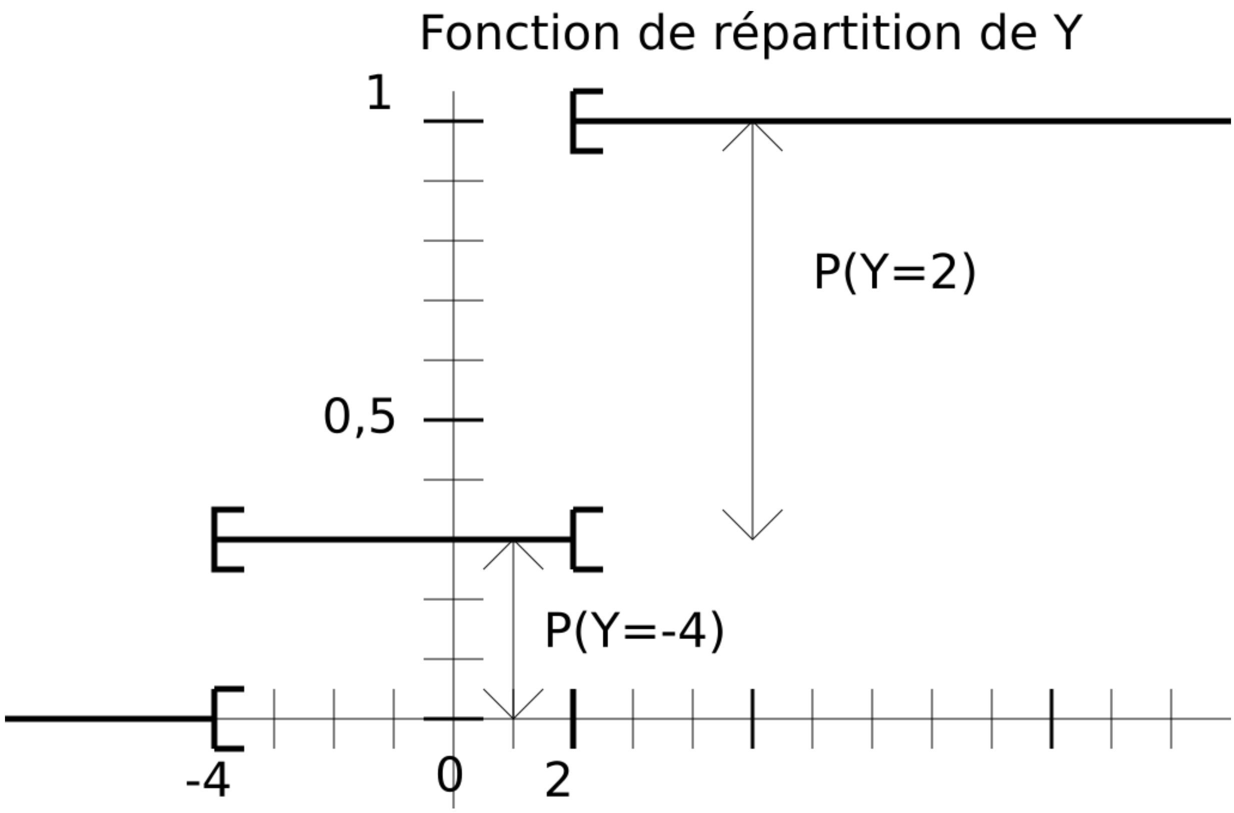
De la variable aléatoire X de loi suivante :

x	0	20	100
P(X=x)	0,6	0,3	0,1



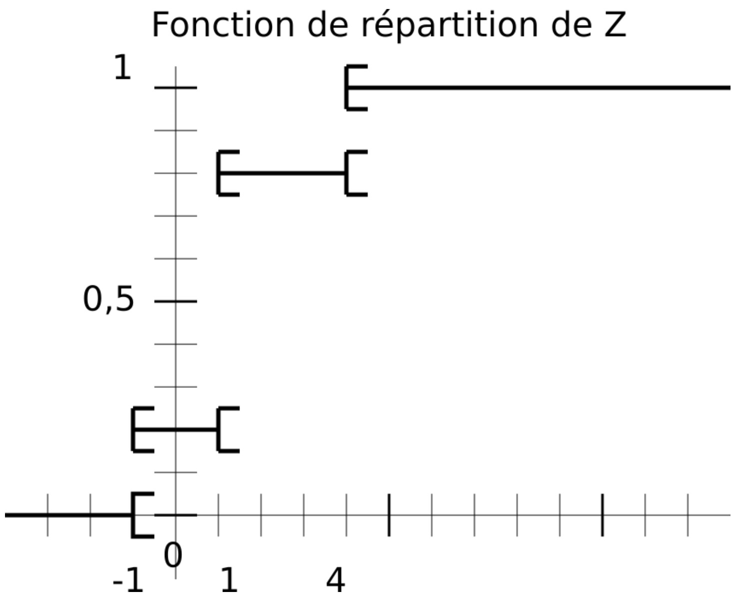
Dessiner le graphe de la fonction de répartition des variables aléatoires suivantes :
Y :

x	-4	2
P(Y=x)	0,3	0,7



Z :

x	-1	1	4
P(Z=x)	0,2	0,6	0,2

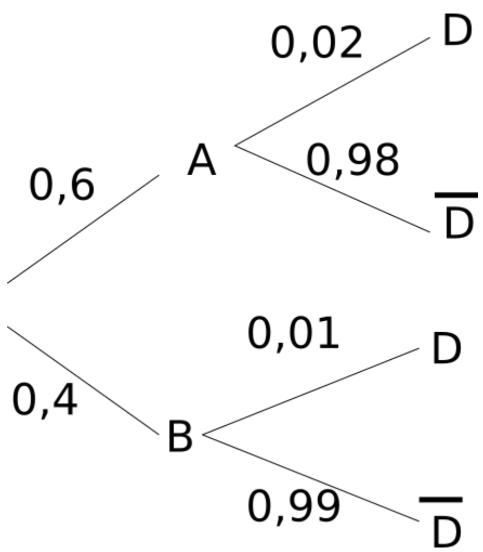


Exercice : Une entreprise fabrique des cartes graphiques pour ordinateurs.
Deux ateliers de fabrication se répartissent la production de la journée.
L'atelier A est plus moderne et produit 60% des cartes et l'atelier B en produit 40%.
Un contrôle de qualité a permis de constater que 2% des cartes de l'atelier A et 1% des cartes de l'atelier B présentent un défaut.
Chaque carte a un coût de production de 10 euros dans l'atelier A et de 15 euros dans l'atelier B.
Les cartes défectueuses sont détruites après contrôle et le coût de recyclage de chaque carte est de 5 euros.
Les autres cartes (celles qui ne sont pas défectueuses) sont vendues 50 euros.

Déterminer le bénéfice que peut espérer en moyenne l'entreprise sur chaque carte fabriquée.

On note A : « la carte est fabriquée dans l'atelier A »
B : « la carte est fabriquée dans l'atelier B »
D : « la carte est défectueuse »
X : le bénéfice

On peut faire l'arbre de probabilité suivant :



D'où $P(A \cap D) = P(X = -15) = 0{,}6 \times 0{,}02 = 0{,}012$

$$P(B \cap D) = P(X = -20) = 0{,}4 \times 0{,}01 = 0{,}004$$

$$P(B \cap \overline{D}) = P(X = 35) = 0{,}4 \times 0{,}99 = 0{,}396$$

$$P(A \cap \overline{D}) = P(X = 40) = 0{,}6 \times 0{,}98 = 0{,}588$$

On en déduit le tableau de la loi du bénéfice :

x	-20	-15	35	40
P(X=x)	0,004	0,012	0,396	0,588

On en déduit que l'espérance vaut

$$E(X) = -20 \times 0{,}004 - 15 \times 0{,}012 + 35 \times 0{,}396 + 40 \times 0{,}588$$

$$E(X) = 37{,}12$$

Le bénéfice moyen est donc de 37,12 euros.

Exercice un peu plus difficile :
Soit une variable aléatoire T telle que

- T a pour valeurs possibles : 1, 2, 3, 4
- P(T = 2) = 2 P(T = 1)
- P(T = 3) = 2 P(T = 2)
- P(T = 4) = 2 P(T = 3)

Déterminer la loi de T et en déduire sa fonction de répartition.

1) Déterminons la loi de T
On sait aussi que
$P(T = 1) + P(T = 2) + P(T = 3) + P(T = 4) = 1$

Posons P(T=1) = a
On a alors P(T=2) = 2P(T=1) = 2a
P(T=3) = 2P(T=2) = 2x2a = 4a
P(T=4) = 2P(T=3) = 2x4a = 8a

D’où a + 2a + 4a + 8a = 1

15a = 1

Finalement a = 1/15

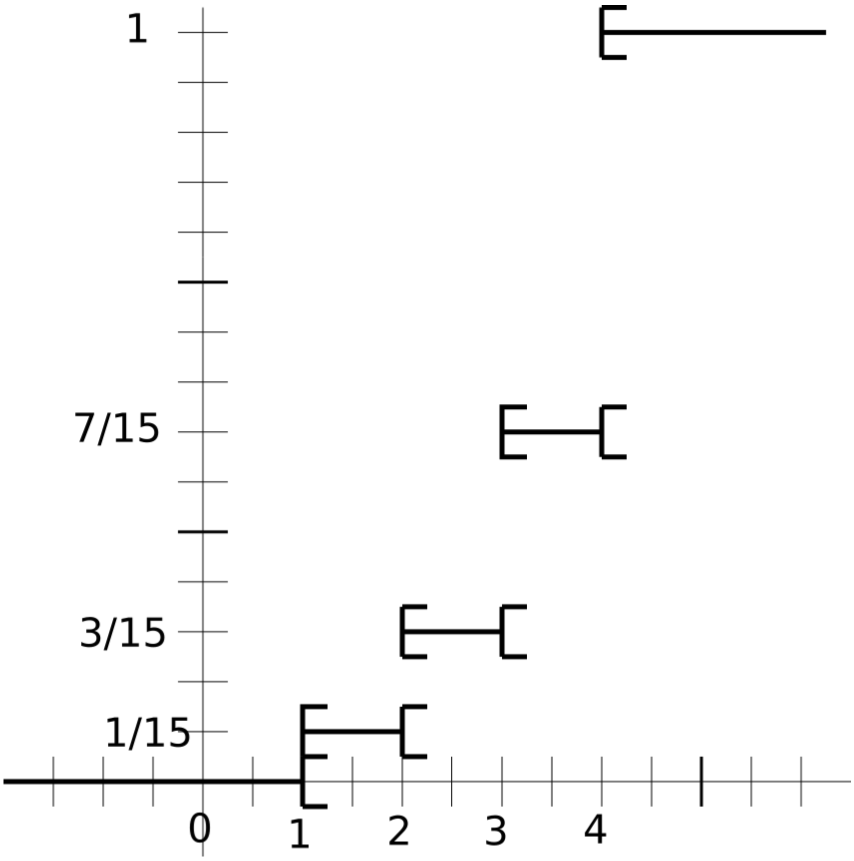
D’où la loi suivante pour T :

x	1	2	3	4
P(T=x)	1/15	2/15	4/15	8/15

2) Fonction de répartition

A vous

x	1	2	3	4
P(T=x)	1/15	2/15	4/15	8/15


1
7/15
3/15
1/15
0
1
2
3
4

V) Test d'indépendance du Khi 2 en pratique

De quoi s'agit-il ?

Nous avons vu précédemment comment vérifier si deux variables sont parfaitement indépendantes à partir d'un tableau croisé.

Dans les exemples, il arrivait que les calculs donnent des résultats « quasi » indépendants,

Le test du chi 2 est une méthode permettant de savoir si on peut accepter l'hypothèse d'indépendance avec un risque (contrôlé) de se tromper.

Exemple : Pour comparer l'efficacité de deux médicaments agissant sur la même maladie, mais aux prix très différents, la Sécurité Sociale a effectué une enquête sur les guérisons obtenues en suivant chacun des traitements. Les résultats sont consignés dans le tableau suivant :

	Médicament cher	Médicament bon marché
Guérisons	54	156
Non-guérisons	6	44

Voici la méthode du test d'indépendance du chi2 par étape :

1) Calcul des totaux :

	Médicament cher	Médicament bon marché	total
Guérisons	$O_{1,1} = 54$	$O_{1,2} = 156$	210
Non-guérisons	$O_{2,1} = 6$	$O_{2,2} = 44$	50
total	60	200	260

2) Calcul des effectifs Théoriques

	Médicament cher	Médicament bon marché	total
Guérisons	$T_{1,1} = \frac{210\times 60}{260} \approx 48{,}462$	$T_{1,2} = \frac{210\times 200}{260} \approx 161{,}538$	210
Non-guérisons	$T_{2,1} = \frac{50\times 60}{260} \approx 11{,}538$	$T_{2,2} = \frac{50\times 200}{260} \approx 38{,}462$	50
total	60	200	260

3) Calcul du chi2calc :

$$\text{Chi2calc}=\sum_{i,j}\frac{(0_{i,j}-T_{i,j})^2}{T_{i,j}}=\frac{(0_{1,1}-T_{1,1})^2}{T_{1,1}}+\frac{(0_{1,2}-T_{1,2})^2}{T_{1,2}}+\frac{(0_{2,1}-T_{2,1})^2}{T_{2,1}}+\frac{(0_{2,2}-T_{2,2})^2}{T_{2,2}}$$

$$\approx\frac{(54-48{,}462)^2}{48{,}462}+\cdots+\frac{(44-38{,}462)^2}{38{,}462}\approx$$

$$0{,}633+0{,}190+2{,}658+0{,}797\approx 4{,}278$$

4) ddl : degré de liberté (ν dans la table du chi2) et VC

l : nombre de lignes du tableau (ne pas compter la ligne Total)

c : nombre de colonnes du tableau (ne pas compter la colonne Total)

ν = ddl = (l-1)x(c-1)

Ici ddl = (2-1)x(2-1) = 1

Valeur critique

Notée V.C., lue dans la table du chi 2, le risque de première espèce α et le nombre de d.d.l=ν étant connus.

En pratique sauf mention du contraire on pendra α=0,05

On lit dans la table pour notre exemple V.C = 3,84

2) Conclusion

Si chi2calc ≥ V.C**, l'hypothèse H_0 d'indépendance des 2 variables est rejetée au risque α**. On conclut que les deux variables ne sont pas indépendantes avec un risque de se tromper égal à **α**.

Si chi2calc < V.C**, l'hypothèse H_0 ne peut être rejetée. (Et donc on l'accepte !)** On conclut que les variables sont indépendantes.

(En tout cas on ne peut pas conclure le contraire !)

Ici on a chi2calc = 4,28 > 3,84=VC donc on rejette l'hypothèse d'indépendance des variables guérison et prix du médicament.

On peut donc raisonnablement estimer ici que le taux de guérison n'est pas indépendant du prix du médicament.

Exercices: A l'aide d'un test du chi2, tester l'indépendance des variables suivantes avec un seuil alpha=0,05 (corrections sur le fichier Excel chi2poly2016)

1) sexe et tabac

	Fument	Ne fument pas
Femme	55	25
Homme	95	25

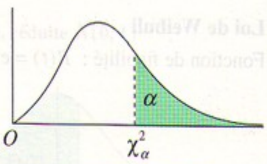
2. couleur des yeux et couleur des cheveux

	Cheveux blonds	Cheveux bruns	Cheveux noirs	Cheveux roux
Yeux bleus	872	380	90	22
Yeux verts ou bruns	500	815	488	33

Table de distribution de χ^2 (loi de K. Pearson)

La table donne la probabilité α, en fonction du nombre de degrés de liberté ν, pour que χ^2 égale ou dépasse une valeur donnée χ^2_α.

$$\alpha = P(\chi^2 \geqslant \chi^2_\alpha)$$



ν	$\alpha = 0{,}990$	$\alpha = 0{,}975$	$\alpha = 0{,}950$	$\alpha = 0{,}900$	$\alpha = 0{,}100$	$\alpha = 0{,}050$	$\alpha = 0{,}025$	$\alpha = 0{,}010$	$\alpha = 0{,}001$
1	0,0002	0,0010	0,0039	0,0158	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,12	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,47	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,13
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,27	19,67	21,92	24,72	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	6,91	7,96	9,31	23,54	26,30	28,84	32,00	39,25
17	6,41	7,56	8,67	10,08	24,77	27,59	30,19	33,41	40,79
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,80	42,31
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	43,82
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	45,32
21	8,90	10,28	11,59	13,24	29,61	32,67	35,48	38,93	46,80
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	48,27
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	49,73
24	10,86	12,40	13,85	15,66	33,20	36,41	39,37	42,98	51,18
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	52,62
26	12,20	13,84	15,38	17,29	35,56	38,88	41,92	45,64	54,05
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	55,48
28	13,57	15,31	16,93	18,94	37,92	41,34	44,46	48,28	56,89
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	58,30
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	59,70

Quand ν est supérieur à 30, on utilise la table de la loi normale (table de l'écart réduit) avec :

$$t = \sqrt{2\chi^2} - \sqrt{2\nu - 1}$$