

1. a.  $X$  est un caractère quantitatif ;  $a < b$  sont des nombres réels et on sait  $f_{X \in [a,b]} > 0.1$ . Quelles sont les affirmations vraies parmi les affirmations suivantes :

- (1) “La fréquence de  $X$  est entre  $a$  et  $b$  avec une probabilité supérieure à 0.1” ;
- (2) “Les valeurs prises par  $X$  sont supérieures à 0.1” ;
- (3) “ $X$  prend des valeurs entre  $a$  et  $b$ ” ;
- (4) “au moins 10% des valeurs prises par  $X$  sont inférieures à  $b$ ” ;
- (6) “ $a > 0.1$ ”.

b.  $X, Y$  sont des caractères quantitatifs ;  $a < b$  et  $c < d$  sont des nombres réels. Interpréter en terme d'évènements les énoncés suivants :

$$f_{X \in [a,b]} = 1, \quad \frac{f_{X \in [a,b] | Y \in [c,d]}}{f_{X \in [a,b]}} \gg 1, \quad f_{X \in [a,b] | Y \in [c,d]} \simeq 1.$$

2. Pour chacun des extraits de texte suivants indiquer la population, les évènements en jeu, la relation entre les probabilités ou fréquences liées à ces évènements :

- “Un an après la sortie de l’université, la probabilité d’être sans emploi est inférieure à 30%. Elle est deux fois plus grande après des études longues.”
- “Le risque de développer la maladie est 20% moins grand chez un patient de plus de 20 ans que chez un patient de moins de 20 ans.”
- “Fumer double le risque de maladie cardio-vasculaire.” pour chacune des deux possibilités suivantes : (1) la population de référence est la population entière ; (2) la population de référence est la population des non-fumeurs.

3. “Fumer double le risque de maladie cardio-vasculaire” (les données sont inventées pour l’exercice)

On se restreint à la population des personnes de plus de 50 ans. On observe que 20% des individus souffrent de maladie cardio-vasculaire et que 70% des malades déclarent être fumeurs. Un sondage au sein de la population des plus de 50 ans indique que 40% d’entre eux sont fumeurs.

a. On note  $M$  l’évènement “être malade” et  $F$  l’évènement “être fumeur”. Calculer les nombres  $\frac{f_{M|F}}{f_M}$  et  $\frac{f_{M|\text{non}F}}{f_{M|\text{non}F}}$ . Quelle population de référence choisira t-on pour affirmer de façon quantifiée le fait que fumer augmente les risques de maladie cardio-vasculaire ? Quel sera le slogan ?

b. Sans tenir compte du sondage, expliquer pourquoi la proportion de fumeur au sein de la population est forcément supérieure à 14%.

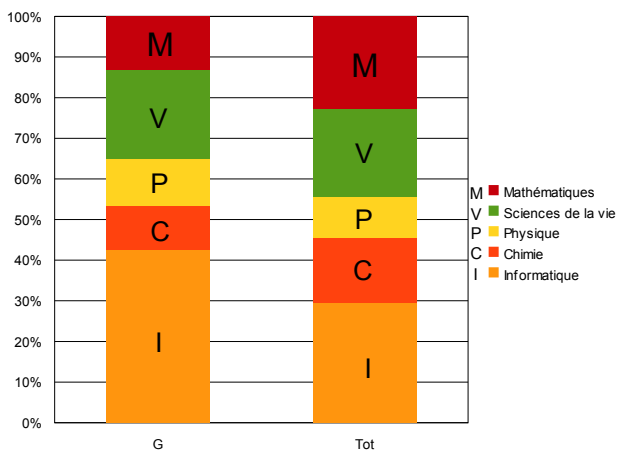
4. Voici ci-dessous un tableau extrait d’une publication de l’Insee

**Tableau 2 : Taux d'activité et recours au temps partiel des hommes et des femmes de 25 à 49 ans selon le nombre et l'âge des enfants**

	Taux d'activité			Temps partiel parmi les personnes ayant un emploi		
	Femmes	Hommes	Ensemble	Femmes	Hommes	Ensemble
Ensemble	83,1	96,2	89,4	29,2	3,7	15,9
Aucun enfant	91,0	94,2	92,8	15,8	5,1	9,6
1 enfant	87,0	97,6	91,5	26,1	3,2	15,4
2 enfants	82,7	97,5	89,3	37,7	2,5	20,2
3 enfants ou plus	63,7	96,6	78,1	47,6	3,3	22,8

- a. Quelle est la population étudiée ? Quels sont les caractères retenus dans le tableau ? Que représentent les nombres : des effectifs conjoints ? des fréquences conjoints, marginales, conditionnelles ? (si oui, de quoi ?) ou autre chose ?
- b. Quelle est la proportion de femmes travaillant à temps partiel parmi toutes les femmes ?
- c. Peut-on déduire de la première ligne du tableau qu'il y a plus d'hommes que de femmes, ou l'inverse, au sein de la population étudiée ?

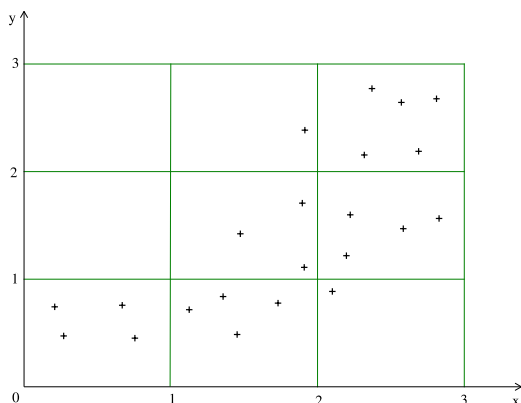
5. (Examen 1ère session mai 2010) On fait une étude statistique des filières (mathématiques, sciences de la vie, etc.) choisies par les étudiants du campus sciences.



Le graphique montre la fréquence des différentes filières parmi les étudiants garçons du campus (première barre) et parmi la population totale du campus (seconde barre).

- a. Quelle est approximativement la fréquence de la filière "Chimie" chez les garçons ? Et quelle est approximativement la fréquence de cette filière dans la population totale ?
- b. Quelles sont les filières dont on observe pratiquement l'indépendance par rapport au sexe de l'étudiant ? Expliquez.

6. (Examen 1ère session avril 2012) Une population de 22 individus est étudiée via deux caractères quantitatifs  $X$  et  $Y$ . Le dessin ci-dessus montre le nuage de points  $(X(s), Y(s))$  associé.



- a. Quel est approximativement l'étendue du caractère  $Y$  ?
- b. On découpe les étendues des caractères  $X$  et  $Y$  en intervalles  $[0, 1[$ ,  $[1, 2[$ ,  $[2, 3[$ . Donnez le tableau des effectifs conjoints des intervalles de  $X$  et  $Y$ .
- c. Quelle est la fréquence de l'évènement " $X \in [0, 1[$ " ?  
Quelle est la fréquence de l'évènement " $X \in [0, 1[$ " conditionné à l'évènement " $Y \in [0, 1[$ " ?
- d. Peut-on affirmer que les caractères  $X$  et  $Y$  sont liés (*i.e.* ne sont pas pratiquement indépendants) ? Justifiez votre réponse.