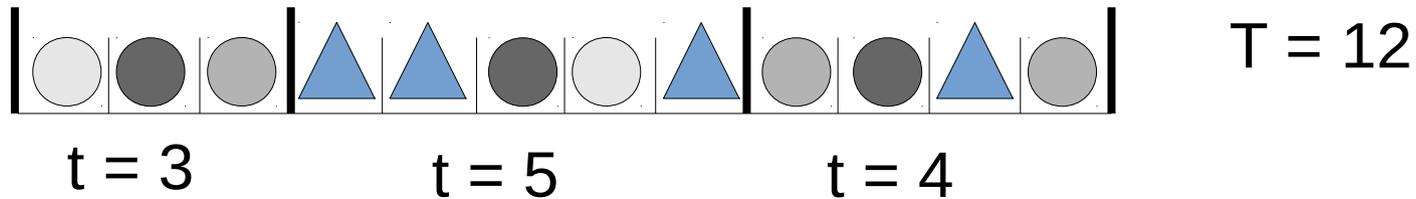


# Introduction illustrée au modèle textométrique des spécificités



Exemple d'un corpus avec 3 parties (par exemple 3 textes),  
et 4 mots différents :



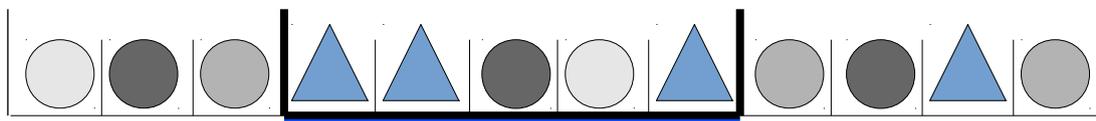
Intéressons-nous à la fréquence du mot « triangle bleu » dans la partie du milieu :

$T = 12$  *taille totale du corpus*

$t = 5$  *taille de la partie*

$F = 4$  *fréquence totale du mot*

$f = 3$  *fréq. du mot ds la partie*

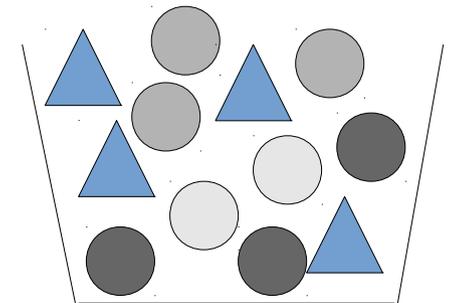


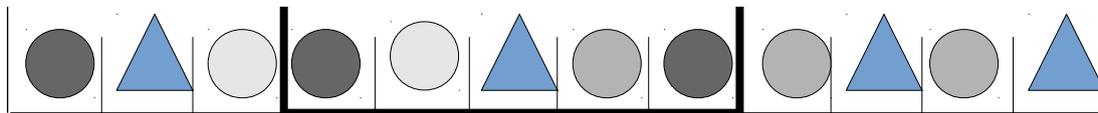
$T = 12$

$t = 5$

$F = 4$

$f = ?$

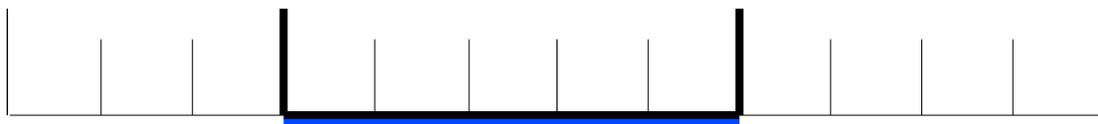
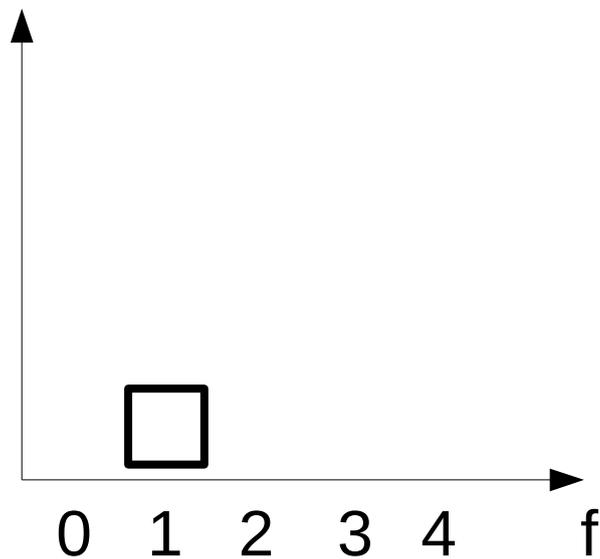




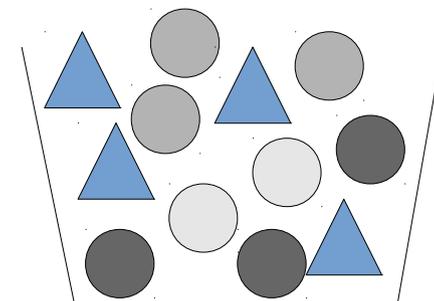
$T = 12$   
 $t = 5$   
 $F = 4$   
 $f = 1$

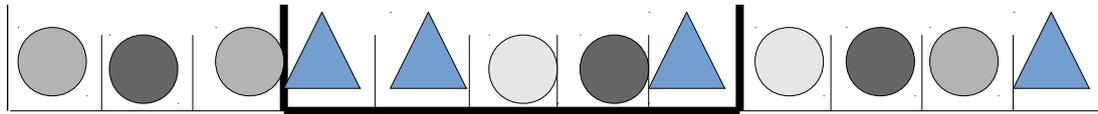


réalisations

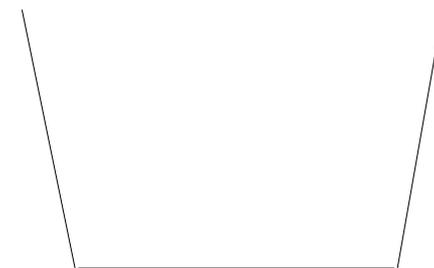


$T = 12$   
 $t = 5$   
 $F = 4$   
 $f = ?$

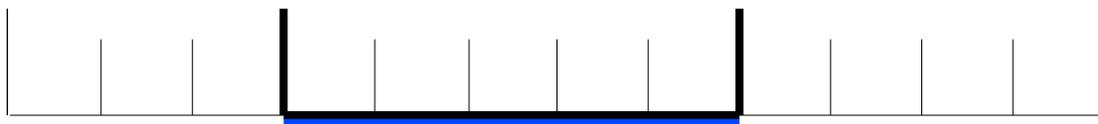
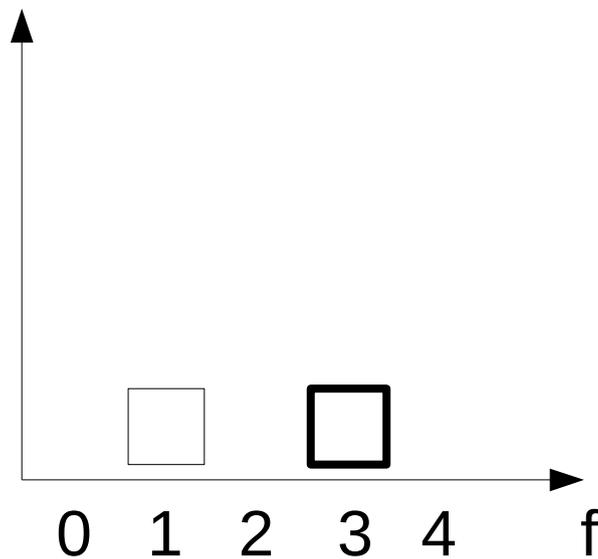




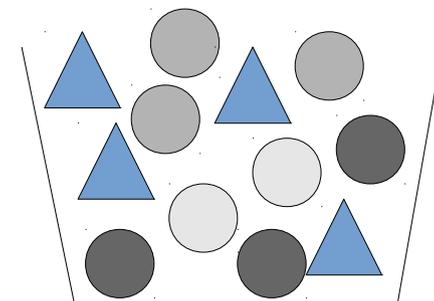
$T = 12$   
 $t = 5$   
 $F = 4$   
 $f = 3$

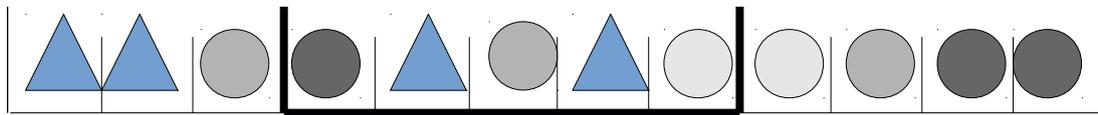


réalisations

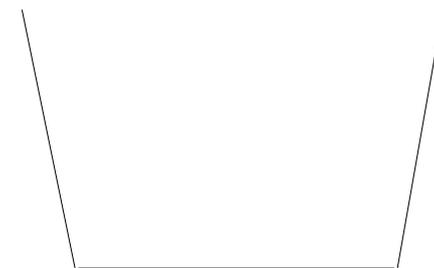


$T = 12$   
 $t = 5$   
 $F = 4$   
 $f = ?$

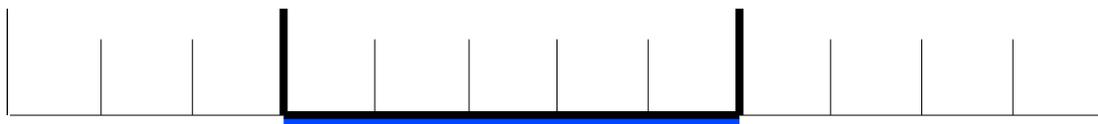
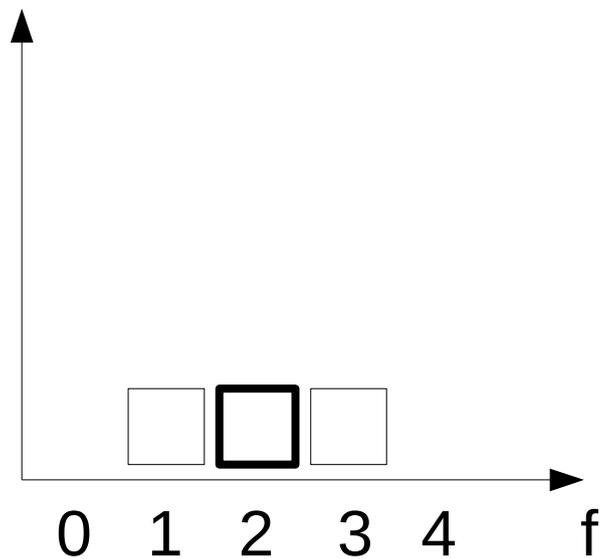




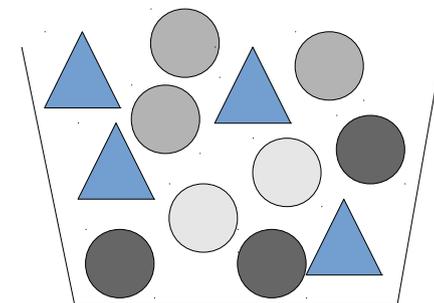
$T = 12$   
 $t = 5$   
 $F = 4$   
 $f = 2$

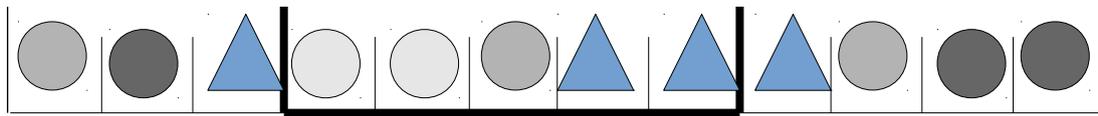


réalisations

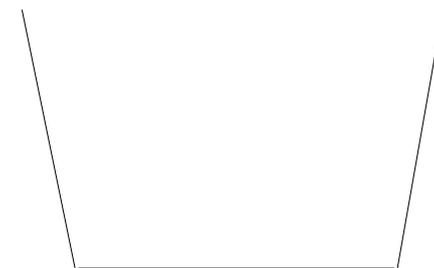


$T = 12$   
 $t = 5$   
 $F = 4$   
 $f = ?$

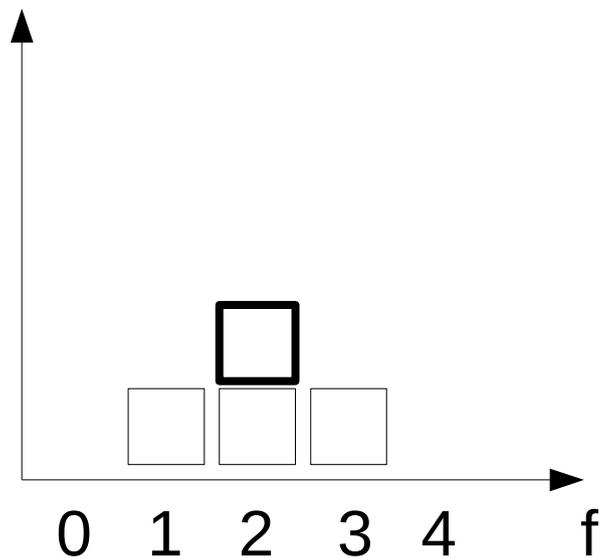




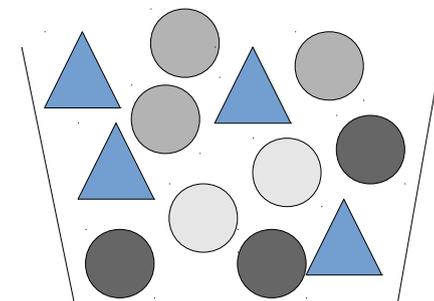
$T = 12$   
 $t = 5$   
 $F = 4$   
 $f = 2$

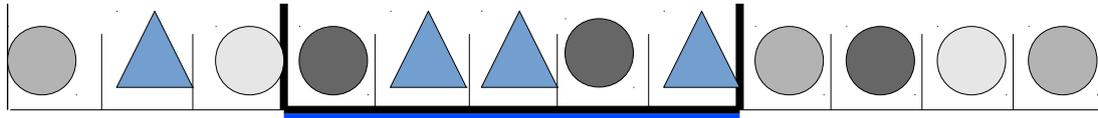


réalisations

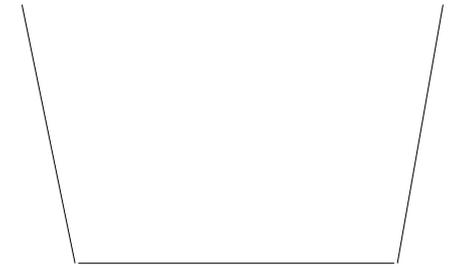


$T = 12$   
 $t = 5$   
 $F = 4$   
 $f = ?$

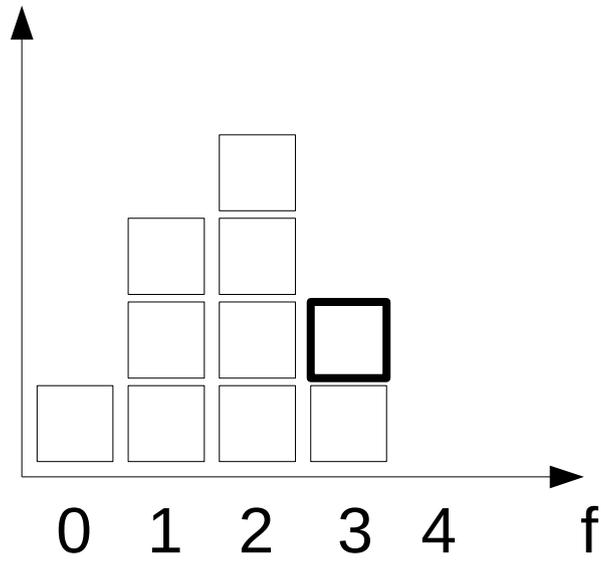


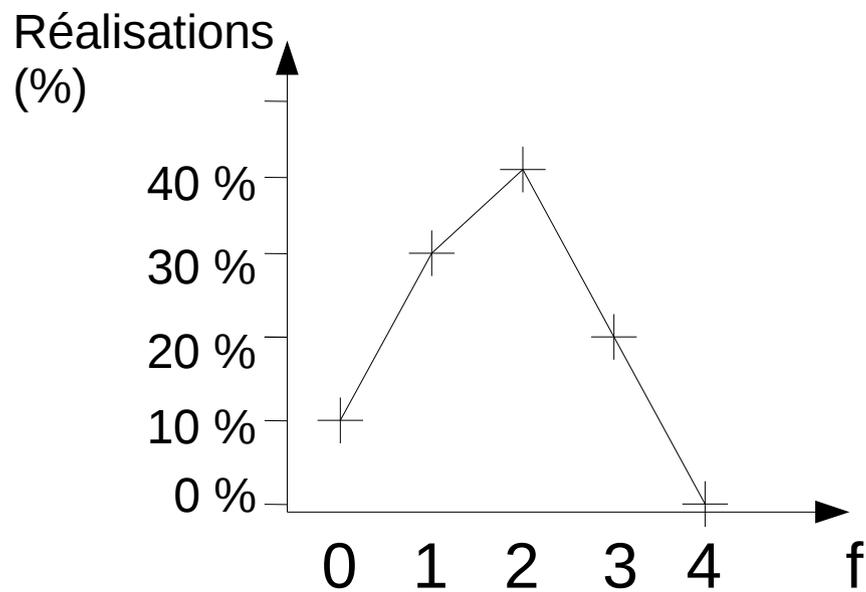
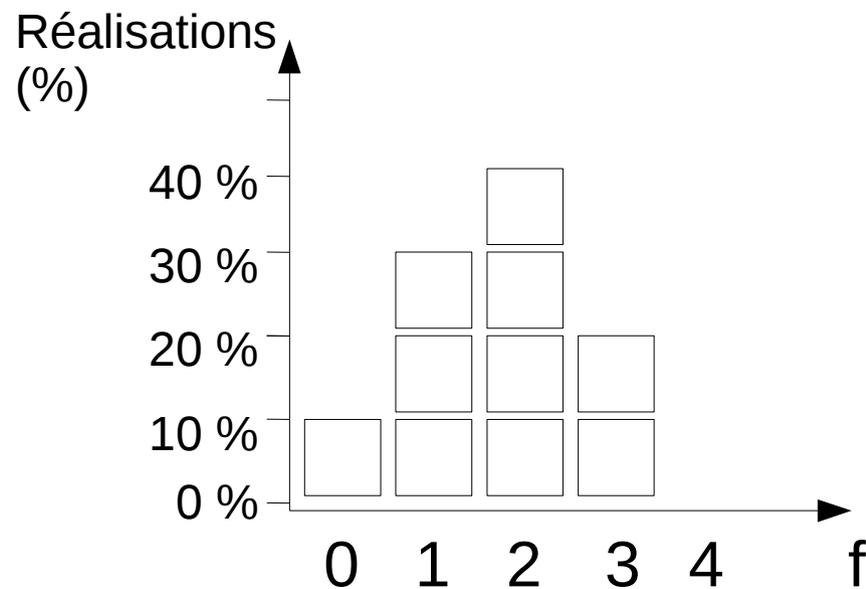
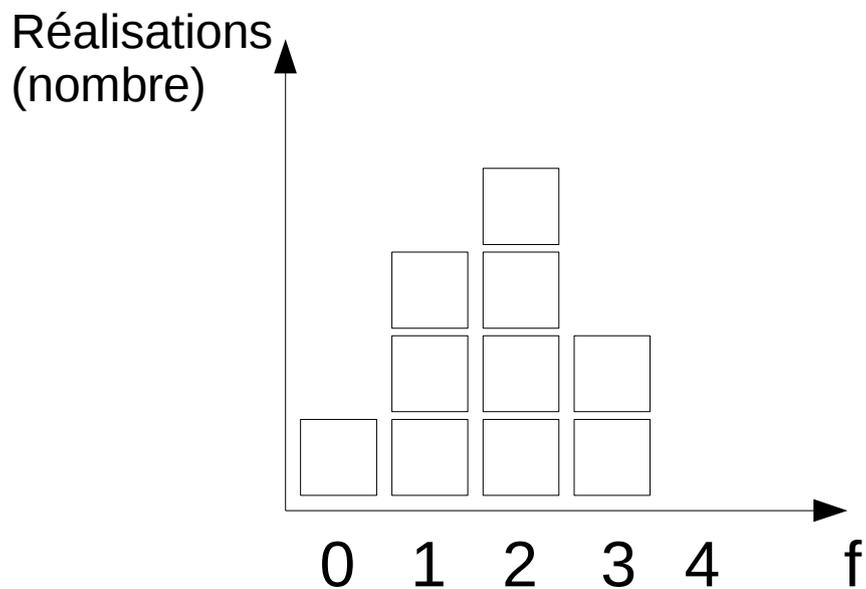


$T = 12$   
 $t = 5$   
 $F = 4$   
 $f = 3$

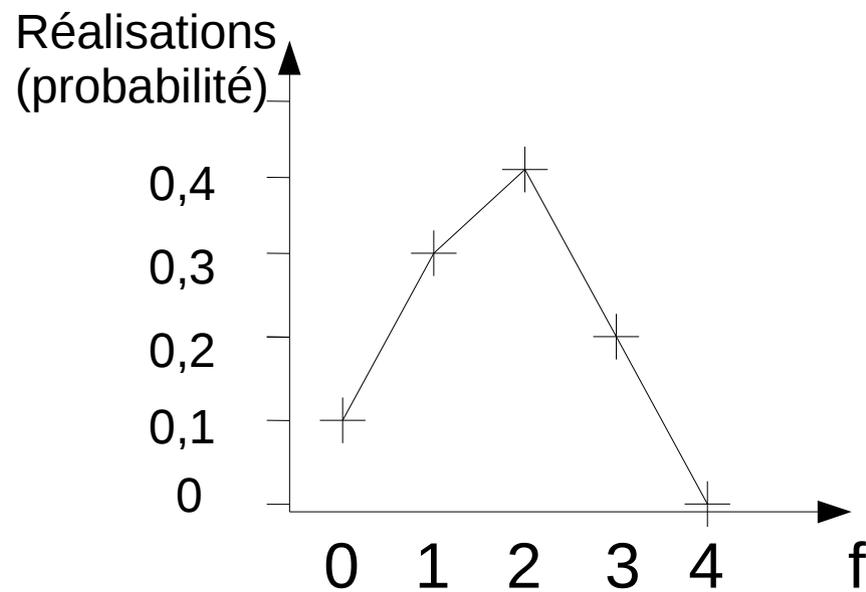


réalisations





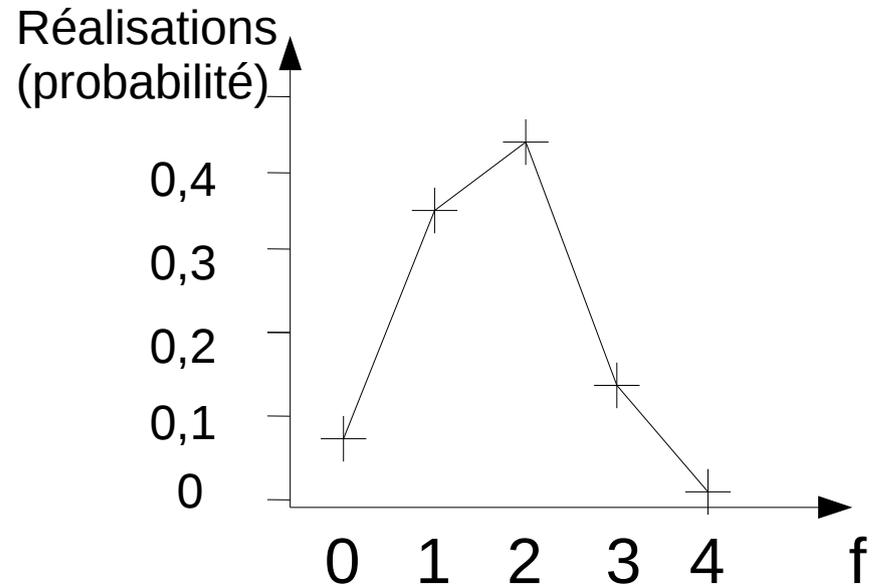
La somme des pourcentages vaut 100.



La somme des probabilités vaut 1.

Après un plus grand nombre d'observations, la position des points se stabilise et se précise :

f	p
0	0,07
1	0,35
2	0,43
3	0,14
4	0,01



Concrètement, on n'aura pas besoin de simuler ces lancers de mots pour trouver leur probabilité, car on a des formules mathématiques pour les calculer directement à partir des quatre paramètres T, t, F, f.

La spécificité du mot « triangle bleu » de fréquence  $f=3$  dans la partie considérée se détermine alors comme ceci :

1) pour son signe : La fréquence attendue comme la plus probable aurait été  $f=2$ , comme  $3 \geq 2$  alors la spécificité est positive.

2) pour sa valeur :

$$\text{probabilité } (f \geq 3) = \text{probabilité } (f = 3) + \text{probabilité } (f = 4) = 0,15 = 1,5 \times 10^{-1} = 10^{-0,8}$$

$$\text{Spécificité (de triangle bleu, dans partie centrale)} = + |\text{Log } (0,15)| = + |\text{Log } (10^{-0,8})| = +0,8$$