Chapitre 4 : Statistiques et probabilités

Julien Reichert

1 Statistiques

Cette section est essentiellement une présentation du vocabulaire général des statistiques, en vue d'un approfondissement au cours des années ultérieures.

Un caractère sur lequel les statistiques se fondent peut être n'importe quelle propriété, pouvant prendre au moins deux valeurs afin que les statistiques aient de l'intérêt. Par exemple, la matière préférée d'un élève est un caractère, la couleur des yeux d'un individu aussi, de même que son âge.

Remarque : Le nombre de valeurs possibles n'est pas nécessairement fini, et lorsqu'il est infini on peut regrouper ces valeurs afin d'avoir un nombre fini de catégories.

Les caractères étudiés seront soit qualitatifs, c'est-à-dire que les valeurs ne sont pas censées être vues comme des nombres ordonnés [1], soit quantitatifs, c'est-à-dire que les valeurs sont nécessairement toutes des réels et elles sont de préférence triées par ordre croissant avant tout calcul.

Dans le cas d'un caractère qualitatif, de nombreuses notions qui s'appliqueraient aux caractères quantitatifs n'ont aucun sens, c'est pourquoi les caractères qualitatifs seront présentés en premier.

Définition
Soit X un caractère qualitatif pouvant prendre des valeurs notées $x_1, x_2, \ldots, x_n$. Sur un échantillon de k personnes, où k est appelé **effectif total**, l'**effectif** d'une valeur x_i, où i est n'importe quel entier entre 1 et n, est le nombre noté k_i de personnes pour lesquelles le caractère prend la valeur x_i. La **fréquence** d'une valeur x_i, notée f_i, est le rapport $\frac{k_i}{k}$.

La somme des effectifs de toutes les valeurs doit nécessairement être k, et la somme de leurs fréquences doit nécessairement être 1, en prenant garde d'additionner les valeurs exactes des fréquences et non des valeurs approchées intermédiaires. [2]

Remarque : Il est parfois utile de ne pas simplifier les fractions au cours des calculs, mais la forme simplifiée reste souhaitée en toute fin.

Exemple : Le tableau ci-dessous donne la couleur des yeux d'un échantillon de 24 personnes. Bien entendu, les nuances de couleur étant infinies, on regroupe les valeurs suivant une nuance dominante.

Couleur	bleu	brun	vert	gris	noir	
Effectif	4	8	1	3	8	24
Fréquence	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{3}$	

Dans cet exemple, les valeurs sont donc x_1 = « bleu », x_2 = « brun », x_3 = « vert », x_4 = « gris » et x_5 = « noir ».
Les effectifs correspondants sont $k_1 = 4$, $k_2 = 8$, $k_3 = 1$, $k_4 = 3$ et $k_5 = 8$.
Les fréquences correspondantes sont $k_1 = \frac{4}{24} = \frac{1}{6}$, $k_2 = \frac{8}{24} = \frac{1}{3}$, $k_3 = \frac{1}{24}$, $k_4 = \frac{3}{24} = \frac{1}{8}$ et $k_5 = \frac{8}{24} = \frac{1}{3}$.

1. Même si les valeurs sont parfois des nombres ; le nombre préféré n'a par exemple pas vocation à être comparé pour deux personnes différentes.
2. Voilà pourquoi dans certains sondages la somme est parfois de 99 % ou 101 %.

Définition
Soit X un caractère quantitatif pouvant prendre des valeurs, nécessairement données dans l'ordre strictement croissant, $x_1 < x_2 < \cdots < x_n$. Les notions d'effectif total, d'effectif pour une valeur et de fréquence pour une valeur sont les mêmes que pour un caractère qualitatif. L'**effectif cumulé croissant** d'une valeur x_i est la somme des effectifs des valeurs de x_1 à x_i incluses. Ainsi, l'effectif cumulé croissant de la première valeur est son effectif et l'effectif cumulé croissant de la dernière valeur est l'effectif total. La **fréquence cumulée croissante** d'une valeur x_i est la somme des fréquences des valeurs de x_1 à x_i incluses. Il s'agit donc aussi du rapport entre l'effectif cumulé croissant de la valeur et l'effectif total. Cette dernière caractérisation facilite d'ailleurs le calcul.

Exemple : Le tableau ci-dessous donne le nombre d'étages par immeuble dans un quartier qui comporte 30 bâtiments.

Nombre d'étages	3	4	5	8	20	
Effectif	5	12	9	3	1	30
Fréquence	$\frac{1}{6}$	$\frac{2}{5}$	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{1}{30}$	
Effectif cumulé croissant	5	17	26	29	30	
Fréquence cumulée croissante	$\frac{1}{6}$	$\frac{17}{30}$	$\frac{13}{15}$	$\frac{29}{30}$	1	

Remarques :

- L'effectif cumulé croissant d'une valeur x_i est le nombre de fois dans l'échantillon où la valeur est inférieure ou égale à x_i.
- Rien n'interdit d'ajouter d'autres valeurs qui n'apparaissent pas, auquel cas elles ont l'effectif 0. En revanche, il faut les insérer au bon endroit et donc procéder à une renumérotation. Quoi qu'il en soit, la lisibilité est affectée.

Définition
Soit X un caractère quantitatif prenant les valeurs $x_1, x_2, \ldots, x_n$. La **moyenne simple** des n valeurs du caractère vaut $\frac{x_1+x_2+\cdots+x_n}{n}$. Il s'agit de la somme des valeurs divisée par le nombre de valeurs.

Exemple : La moyenne simple du nombre d'étages par immeuble, sans tenir compte du nombre d'immeubles ayant un nombre d'étages donné, est $\frac{3+4+5+8+20}{5} = 8$.

Définition
Soit X un caractère quantitatif prenant les valeurs $x_1, x_2, \ldots, x_n$. La **moyenne pondérée** du caractère sur un échantillon de k personnes est donnée par la formule

$$\bar{x} = \frac{k_1x_1 + k_2x_2 + \cdots + k_nx_n}{k},$$

où k_1, k_2, $\ldots$, k_n sont les effectifs correspondant aux valeurs de même indice. La moyenne pondérée est donc la somme des valeurs comptées avec leurs multiplicités divisée par l'effectif total.

Exemple : La moyenne pondérée du nombre d'étages par immeuble est $\frac{5\times3+12\times4+9\times5+3\times8+1\times20}{30} = \frac{152}{30}$ soit un peu plus de 5.

La moyenne pondérée est potentiellement très éloignée de la moyenne simple, il suffit d'imaginer qu'au lieu de 20 étages, le plus grand immeuble en ait eu 100. La moyenne simple aurait été de 24 alors que la moyenne pondérée aurait encore été inférieure à 8. Comme son nom l'indique, la moyenne pondérée donne plus de poids à certaines valeurs, comme les coefficients donnent plus de poids à certaines notes.

Définition
Soit X un caractère quantitatif prenant les valeurs $x_1, x_2, \ldots, x_n$ avec les effectifs respectifs $k_1, k_2, \ldots, k_n$. La médiane est un réel tel que la moitié des valeurs prises soient inférieures à lui et la moitié des valeurs prises soient supérieures à lui. Par convention, on décide que si l'effectif total est impair, alors la médiane est la « valeur du milieu », en n'oubliant pas de compter chaque valeur avec son effectif, et si l'effectif total est pair, alors la médiane est la moyenne simple entre les deux valeurs du milieu, qui ne sera alors pas nécessairement une valeur du caractère.

Exemple : La médiane du nombre d'étages par immeuble est 4, car parmi les 30 immeubles, 5 ont 3 étages, on prend 9 sur les 12 immeubles de 4 étages et on considère que ce sont les 9 valeurs suivantes, de l'autre côté, 13 ont 5, 8 ou 20 étages et on leur adjoint un immeuble de 4 étages. Les deux immeubles du milieu sont les deux immeubles à 4 étages restants, et la moyenne de 4 et 4 est bien entendu 4.

Plus généralement, pour un effectif total impair de k, la médiane est la $\frac{k+1}{2}$-ième valeur (toujours comptée avec l'effectif), et pour un effectif total pair de k, la médiane est la moyenne entre la $\frac{k}{2}$-ième et la $(\frac{k}{2}+1)$-ième valeur. L'effectif cumulé croissant permet de localiser la valeur en question.

Définition
Soit X un caractère quantitatif prenant les valeurs $x_1, x_2, \ldots, x_n$ avec les effectifs respectifs $k_1, k_2, \ldots, k_n$. Le **premier quartile** est la plus petite valeur du caractère, notée Q_1, telle qu'au moins le quart, soit 25 %, des valeurs prises soient inférieures ou égales à lui [3]. Le **troisième quartile** est la plus petite valeur du caractère, notée Q_3, telle qu'au moins trois quarts, soit 75 %, des valeurs prises soient inférieures ou égales à lui [4].

Pour déterminer les quartiles, on arrondit $\frac{k}{4}$ et $\frac{3k}{4}$ à l'entier supérieur (donc si c'est déjà un entier, on ne le change pas), ce qui donnera le rang respectif de Q_1 et Q_3. L'effectif cumulé croissant permet là aussi de localiser la valeur en question.

Exemple : Le premier quartile du nombre d'étages par immeuble est 4 et le troisième quartile est 5. On a pris respectivement la 8^{e} valeur et la 23^{e} valeur.

Définition
Soit X un caractère quantitatif prenant les valeurs $x_1, x_2, \ldots, x_n$, les effectifs n'important pas. L'**étendue** est la différence entre la plus grande et la plus petite valeur, soit $x_n - x_1$.

Définition
Soit X un caractère quantitatif. L'**écart interquartile** est le réel $Q_3 - Q_1$.

Ces deux réels sont des critères de dispersion, au contraire de la moyenne et de la médiane par exemple, qui sont des critères de position. La médiane, les quartiles et donc l'écart entre eux sont peu sensibles aux valeurs extrêmes et à des modifications de celles-ci, au contraire de l'étendue, ainsi que de la moyenne quand les modifications sont importantes au regard de l'effectif.

Outre les tableaux, on peut utiliser des diagrammes comme l'histogramme (donnant l'effectif, voire la fréquence) et le diagramme circulaire (donnant la fréquence, en précisant éventuellement l'effectif), pour un caractère aussi bien qualitatif que quantitatif. En revanche, le nuage de points est réservé aux caractères quantitatifs. Ces nuages de points prennent pour abscisses les valeurs et pour ordonnées, suivant un choix fait à l'avance, soit l'effectif de la valeur en abscisse, soit son effectif cumulé croissant.

Définition
Effectuer un **échantillonnage** de taille k, c'est sélectionner au hasard et indépendamment k éléments d'une population et relever leur valeur pour un certain caractère, qui peut être qualitatif comme quantitatif.

Exemple : On interroge $k = 100$ personnes dans la rue et on leur demande leur pointure. On choisit $k = 10$ stylos dans une boîte qui en contient des centaines et on relève leur couleur.

Remarque : La notion de hasard et d'indépendance est assez flexible en statistiques : on considèrera que le choix est indépendant alors qu'on cherchera souvent tout de même à ne pas avoir deux fois le même individu. De même, les sondages se font sur ce qui est qualifié d'échantillon représentatif, où les instituts s'arrangent pour contrôler le nombre de sondés d'un certain âge, ou d'un certain sexe, etc.

Définition
Une **épreuve de Bernoulli** est un test qui n'a que deux issues. Par habitude, on les note « vrai » et « faux » ou 0 et 1 (afin d'avoir un caractère quantitatif).

---

3. et 75 % des valeurs prises lui sont alors supérieures ou égales
4. et 25 % des valeurs prises lui sont alors supérieures ou égales

Exemple : L'épreuve de Bernoulli la plus classique est un lancer de pièce. Cependant, rien n'oblige les deux issues à être équiprobables. Le loto est aussi une épreuve de Bernoulli : soit on gagne, soit on perd, et même soit on gagne au plus haut rang, soit ce n'est pas le cas. Dans cet exemple le plus extrême, l'issue favorable a une probabilité sous le millionième.

Proposition
Soit une épreuve de Bernoulli telle que la probabilité de l'issue retenue comme favorable soit entre $0,2$ et $0,8$ (et donc la probabilité de l'autre issue est aussi entre $0,2$ et $0,8$). Si on fait un échantillonnage de k résultats de cette épreuve, où k est assez grand (on considère qu'il doit être supérieur à 25), la fréquence de l'issue favorable sera dans l'intervalle $[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}]$ avec une probabilité supérieure à 95 %.

Remarque : L'utilité de cet intervalle, nommé **intervalle de fluctuation** au seuil de 95 %, sera donnée dans les classes supérieures. Il s'agit en quelque sorte de deviner une valeur pour la probabilité p et de vérifier la cohérence de cette hypothèse en regardant si la fréquence de suffisamment parmi de nombreux échantillonnages successifs de k résultats est dans l'intervalle déduit.

2 Probabilités

Cette section reprend le vocabulaire de base sur les ensembles, qui est donc à revoir avant tout.

Il est important de garder à l'esprit que, quoique les notions de statistiques et de probabilités soient liées, une probabilité ne donne qu'une tendance sur la statistique et une échantillonnage ne permet pas de déduire une probabilité. Il reste vrai que si un sac contient trois boules rouges, deux boules noires et cinq boules blanches, en piochant une seule boule, elle aura une chance sur deux d'être blanche : ici, il ne s'agit pas d'un échantillonnage mais d'une liste exhaustive des possibilités.

Définition
On considère une expérience aléatoire, qui peut aller du lancer d'une pièce à un mélange de cartes en passant par le tirage du loto. Une **issue** est un résultat possible de l'expérience aléatoire, par exemple « pile », une permutation des cartes ou une combinaison de boules entre 1 et 50 plus un numéro complémentaire. L'ensemble des issues est appelé l'**univers** associé à l'expérience et noté Ω. Il est important de savoir calculer le nombre de ses éléments. Un **événement** A est un sous-ensemble (aussi appelé une partie) de l'univers, donc un ensemble d'issues. On dit qu'une issue **réalise** un événement si elle appartient à l'ensemble correspondant. Des événements correspondant aux expériences sont par exemple $\{\text{face}\}$, l'ensemble des mélanges tels que la troisième carte soit un pique ou l'ensemble des tirages figurant sur une grille de loto qu'on a jouée. L'événement **impossible**, noté $\varnothing$, correspond à l'ensemble vide ; l'événement **certain** est Ω et un événement **élémentaire** est un sous-ensemble de Ω de taille 1.

Les notions d'intersection, de réunion et de complémentaire (ou contraire) d'événements sont les mêmes que pour les ensembles.

Ainsi, dans le cas du lancer d'un dé à 6 faces, si on note A l'événement « obtenir un nombre pair », soit l'ensemble $\{2; 4; 6\}$, et B l'événement « obtenir un multiple de 3 », soit $\{3; 6\}$, l'intersection des événements A et B est l'événement « obtenir un nombre multiple de 2 ET de 3 », soit l'ensemble $\{6\}$, qui correspond donc à un événement élémentaire, la réunion des événements A et B est l'événément « obtenir un nombre multiple de 2 OU de 3 », soit l'ensemble $\{2; 3; 4; 6\}$, et le contraire de l'événement A, noté $\bar{A}$, est l'événément « obtenir un nombre impair », soit l'ensemble $\{1; 3; 5\}$.

Définition
Deux événéments sont dits **disjoints** ou **incompatibles** lorsque leur intersection est l'événement impossible. Les événements A et $\bar{A}$ sont toujours disjoints, par définition.

Définition
On considère une expérience aléatoire, avec le vocabulaire défini précédemment. Une **loi de probabilité** sur l'univers (fini) Ω associe à chaque issue une probabilité, qui est un réel positif (potentiellement nul, forcément inférieur ou égal à 1 d'après ce qui suit), de sorte que la somme des probabilités de toutes les issues soit exactement 1.

Dans le cas où les issues sont notées $e_1, e_2, \ldots, e_n$, on peut noter p_i la probabilité de chaque issue e_i, appelée probabilité de l'événement élémentaire $\{e_i\}$. **La probabilité d'un événement est la somme des probabilités des issues qui le composent.** On note $p(A)$ la probabilité de l'événément A.

Définition
Une **situation d'équiprobabilité** signifie que toutes les issues de l'univers ont la même probabilité. Quand l'univers est fini et comporte n issues, la probabilité de chacune est alors $\frac{1}{n}$. La loi de probabilité associée est aussi appelée **loi uniforme**.

Proposition
En situation d'équiprobabilité, la probabilité d'un événement A est alors donnée par la formule $\frac{\text{nombre d'issues réalisant } A}{\text{nombre total d'issues}}$.

Il est généralement recommandé de préciser au début d'un exercice qu'on est en situation d'équiprobabilité, du moins lorsque c'est le cas. Cette situation est très fréquente dans les modèles mathématiques, elle se reconnaît à des mots-clés comme « dé non pipé », « pièce équilibrée », « on tire au hasard une boule ». Attention : dans le dernier cas, si l'expérience consiste à regarder la couleur de la boule, bien qu'il y ait équiprobabilité sur les boules, suivant le nombre de boules de chaque couleur la probabilité diffèrera.

Proposition
Les formules basiques données ci-après sont toujours valables, même hors situation d'équiprobabilité, et elles seraient même valables si l'univers était infini.
- Soit A un événement. Alors $p(\bar{A}) + p(A) = 1$. En effet, toute issue réalise exactement un des événements A ou $\bar{A}$.
- Soient A et B deux événements. Alors $p(A \cup B) = p(A) + p(B) - p(A \cap B)$. En particulier, si A et B sont incompatibles, la formule devient $p(A \cup B) = p(A) + p(B)$ car on retire la probabilité de l'événement impossible. En fait, pour trouver la probabilité de $A \cup B$, on additionne les probabilités de toutes les issues de A et de celles de B, mais celles de $A \cap B$ sont alors comptées une fois de trop et il faut les retirer.

Pour revenir sur la remarque faite en début de section, la probabilité donne une tendance pour la statistique dans la mesure où la fréquence de chaque issue se rapproche de sa probabilité d'autant plus que le nombre d'expériences devient grand. Cependant, jouer au loto n fois ne suffit pas pour gagner au moins une fois (et ce n'est pas rentable non plus au total), en notant $\frac{1}{n}$ la probabilité, car la notion de « grand » ici se réfère à approcher l'infini. Cette propriété est appelée la loi des grands nombres. Ceci se retrouve en prenant l'intervalle de fluctuation au seuil de $95\,\%$ et en regardant la valeur de $\frac{1}{\sqrt{N}}$ quand N augmente.

3 Compléments : probabilités (un peu plus) avancées

Cette section est une introduction aux dénombrements et aux probabilités de niveau Terminale. Elle a pour seul but de satisfaire la curiosité intellectuelle ou de fournir le bagage permettant de jouer correctement aux jeux de cartes, entre autres.

3.1 Bases du dénombrement

Soit un ensemble E à 6 éléments distincts. On souhaite calculer le nombre de façons d'ordonner E. Le nombre de choix possibles pour le premier élément est 6 : tous les éléments de E conviennent. Après cela, le nombre de choix possibles pour le deuxième élément n'est plus que 5, car l'élément retenu pour être le premier n'est plus disponible. Ensuite, le nombre de choix possibles pour le troisième élément est de 4 pour la même raison, et ainsi de suite. Le nombre de façons d'ordonner E, ou nombre de permutations, est alors $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$.

Dans le cas d'un ensemble à n éléments, ce principe s'étend et le nombre de permutations est de $n \times (n-1) \times \cdots \times 2 \times 1$. On note $n!$ et on appelle « factorielle n » ce nombre.

Reprenons l'ensemble E, on souhaite calculer le nombre de façons de prendre 2 éléments de E en leur donnant un ordre. Cette fois-ci, le calcul s'arrête une fois le deuxième élément sélectionné. Il y a donc $6 \times 5 = 30$ possibilités. En quelque sorte, on a fait $6 \times 5 \times 4 \times 3 \times 2 \times 1$ qu'on a divisé par ce qui n'était pas nécessaire, à savoir $4 \times 3 \times 2 \times 1$.

Dans le cas d'un ensemble à n éléments, le nombre de façons de prendre k éléments de E en leur donnant un ordre, aussi appelé nombre d'arrangements de k parmi n éléments, est $n \times (n-1) \times \cdots \times (n-k+1)$. Il y a k facteurs dans ce produit, qu'on écrit plus simplement $\frac{n!}{(n-k)!}$.

Toujours avec l'ensemble E, on souhaite maintenant calculer le nombre de façons de prendre 2 éléments de E, mais sans ordre cette fois. Les 30 possibilités obtenues avant se rassemblent en 15 paires de façons indistinctibles. En fait, on a divisé 30 par le nombre de façons d'ordonner les 2 éléments retenus afin d'éviter les redondances.

Dans le cas d'un ensemble à n éléments, le nombre de façons de prendre k éléments de E sans ordre, aussi appelé nombre de combinaisons de k parmi n éléments, est $\frac{n \times (n-1) \times \cdots \times (n-k+1)}{k \times (k-1) \times \cdots \times 1}$. Il y a k facteurs au numérateur et k au dénominateur, et on écrit plus simplement ce nombre de combinaisons $\frac{n!}{(n-k)!k!}$. Ce nombre est également appelé coefficient binomial d'indices $(n;k)$ et écrit $\binom{n}{k}$.

3.2 Paradoxe de Von Mises

Ce paradoxe, aussi appelé paradoxe des anniversaires, est classiquement vulgarisé ainsi : « Dans une population donnée, le nombre de personnes suffisant pour que la probabilité qu'au moins deux soient nées le même jour de l'année est 28. »

Le côté paradoxal est le nombre étonnamment bas, alors qu'une réponse intuitive serait de l'ordre de 180. En fait, si on interroge les n personnes de la population au hasard, la première est née un certain jour. La probabilité que la deuxième soit née un jour différent (en négligeant désormais l'influence des années bissextiles) est $\frac{365}{366}$. Ensuite, si c'est le cas, la probabilité que la troisième soit née un jour différent des deux autres est $\frac{364}{366}$. On se retrouve donc à faire un produit de probabilités toujours plus petites, et pour $n = 23$, cela donne

$$1 \times \frac{365}{366} \times \frac{364}{366} \times \frac{363}{366} \times \cdots \times \frac{344}{366}$$

soit environ $0,494$. Pour $n = 22$, la probabilité que tous soient né un jour différent reste un peu supérieure à $0,5$, en revanche.

3.3 Principe du moindre choix

Ce principe est, lui aussi, un classique de la vulgarisation. Il est connu dans le monde anglophone pour avoir été mis en œuvre dans un show télévisé. Le principe est le suivant : un candidat a le choix entre trois rideaux, dont un seulement contient un cadeau de grande valeur. Après un choix préliminaire, l'animateur révèle un rideau contenant un cadeau de petite valeur et le candidat a la possibilité de changer son choix. Son choix préliminaire avait bien entendu une chance sur trois d'être le bon. Cependant, à la deuxième étape, les probabilités ne sont pas changées en $\frac{1}{2}$ partout. Bien au contraire, changer d'avis est gagnant avec deux chances sur trois…en quelque sorte, la probabilité du choix préliminaire reste à $\frac{1}{3}$.

Ceci s'explique en représentant toutes les possibilités : le candidat pouvait avoir choisi le rideau de gauche, du milieu ou de droite, et l'animateur révèle le rideau de gauche, du milieu ou de droite. Cependant, l'animateur ne peut pas révéler le rideau contenant le cadeau de valeur ni le rideau choisi par le candidat. Dans ce cas, si le candidat a choisi un rideau qui n'était pas le bon, l'animateur n'a pas le choix, et le candidat doit changer d'avis.

Ce principe est bien connu des joueurs de bridge qui doivent capturer à la fois la dame et le valet des adversaires dans une couleur.

3.4 Probabilités pour les joueurs de cartes

Cette section présente un calcul de probabilités auxquels un joueur de cartes peut faire face. Le jeu considéré ici [5] a la caractéristique que chaque joueur actif n'aura pas la connaissance de la main de deux autres joueurs. Il s'agit de reconstituer au maximum ces mains en fonction d'informations glanées au fur et à mesure d'une donne.

Si dans une couleur un certain nombre de cartes sont dans les deux mains cachées, comment ces cartes sont-elles réparties, en l'absence totale d'informations ? En théorie, la répartition est la plus centrée possible, mais en pratique la répartition ne sera pas forcément équitable.

Par exemple, il y a 4 cartes cachées, chaque couleur et chaque main disposant de 13 cartes. La répartition « deux cartes dans chaque main » est plus probable que « une carte chez le joueur X et trois chez le joueur Y », mais moins probable que la répartition « une carte dans une main et trois cartes dans l'autre », cumulant deux probabilités.

Le calcul s'effectue ainsi : combien y a-t-il de mains possibles pour le joueur X ? Il s'agit exactement du nombre de répartitions des 26 cartes cachées, 13 dans chaque main. Cela fait donc le nombre de combinaisons de 13 parmi 26, soit $\frac{26!}{13! \times 13!} = 10400600$.

Ensuite, combien y a-t-il de mains du joueur X avec une carte de la couleur ? Il y a 4 choix pour la seule carte de la couleur, et pour les 12 autres cartes de la main, c'est le nombre de façons de prendre 12 cartes parmi 22. Le total est donc $4 \times \frac{22!}{12! \times 10!} = 2586584$. Bien entendu, il y a tout autant de mains du joueur X avec trois cartes de la couleur, car à chaque telle main correspond une main du joueur Y avec une carte de la couleur. La probabilité de la répartition dite 3-1 est donc $\frac{2 \times 2586584}{10400600}$, soit environ $0,497$.

Enfin, pour le nombre de mains du joueur X avec deux cartes de la couleur, il y a 6 choix de paires de cartes de la couleur (c'est le nombre de combinaisons de 2 parmi 4) et on multiplie ce nombre par le nombre de façons de prendre 11 cartes parmi 22. Le total est donc $6 \times \frac{22!}{11! \times 11!} = 4232592$. La probabilité correspondante est alors $\frac{4232592}{10400600}$, soit environ $0,407$.

Le reste correspond aux mains où l'un des joueurs n'a pas de carte de la couleur, donc moins de 5 % pour un joueur donné.

5. le bridge, pour ne pas le nommer