HAI718 Probabilité et statistiques

TD1 : Introduction - Lois usuelles Correction

1 Combinatoire et premiers calculs de probabilités

Exercice 1

1. *En considérant les 26 lettres de l'alphabet, combien peut-on former de mots de 2 lettres ?*
   Pour chacune des deux lettres, on a 26 possibilités, ces possibilités se multiplient, on a donc $26^2 = 676$ mots de deux lettres. Il s'agit d'un tirage avec remise.

   Combien peut-on former de mots de deux lettres constitués d'une consonne suivie d'une voyelle ?
   On a 20 possibilités pour la consonne, 6 pour la voyelle, donc $20 \times 6 = 120$ mots de deux lettres constitués d'une consonne suivie d'une voyelle.

   Combien peut-on former de mots de 2 lettres constitués d'une consonne et d'une voyelle
   On se ramène au problème précédent en remarquant qu'il y a autant de mots de deux lettres constitués d'une voyelle suivie d'une consonne que de mots de deux lettres constitués d'une consonne suivie d'une voyelle, et que ce sont là tous les mots possibles constitués d'une voyelle et d'une consonne. Il y a donc 2×120 mots de deux lettres constitués d'une consonne et d'une voyelle, sans tenir compte de l'ordre (remarque, ici il y avait deux ordres possibles, d'où le facteur 2).
2. *Combien d'équipes différentes de 3 personnes peut-on former à partir d'un groupe de 5 personnes ?*
   On suppose que A, B, C, D et E sont les 5 personnes. On peut énumérer tous les groupes de 3 personnes différentes(sans tenir compte de l'ordre) : ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE. Il y en a 10.
   Pour généraliser au choix de p personnes parmi n : le nombre de combinaisons de p parmi n est $C_n^p = \frac{n!}{p!(n-p)!}$. Dans notre cas, $p = 3$ et $n = 5$, et $C_5^3 = \frac{5\times4\times3\times2\times1}{(3\times2\times1)(2\times1)} = 10$. Il s'agit d'un tirage simultané sans remise.
3. *Avec 17 chevaux au départ, combien y a-t-il de tiercés possibles ?*
   Il y a 17 possibilités pour le premier cheval, 16 pour le second (le premier est déjà arrivé), et 15 pour le troisième, il y a donc $17 \times 16 \times 15 = 4080$ tiercés dans l'ordre.
   Généralisation : le choix **ordonné** de p éléments parmi n est un arrangement, noté $A_n^p = \frac{n!}{(n-p)!}$. Ici $p = 3$ et $n = 17$, et $A_{17}^3 = 4080$. Il s'agit d'un tirage successif sans remise.

   Dans le désordre ?
   On est dans le même cas que le 2)., à savoir le choix de 3 parmi 17 :

il y a donc $C_{17}^3 = 680$ tiercés dans le désordre.
Remarque : $A_n^p = p! \times C_n^p$, et $p!$ est le nombre de façons différentes d'ordonner p éléments.

Exercice 2 *Une urne contient n boules blanches ($n \geq 5$) et 10 boules noires. On tire au hasard et simultanément 10 boules de l'urne.*

1. *Quelle est la probabilité p_n pour que l'on ait tiré exactement 5 boules noires ?*
   $p_n = \dfrac{\text{nombre de tirages avec 5 boules noires}}{\text{nombre de tirages total}}$

   Nombre de tirages total : les 10 boules sont tirées simultanément, cela revient donc à choisir 10 boules parmi $n+10$, il y a donc C_{n+10}^{10} tirages possibles.

   Nombre de tirages avec 5 boules noires : $C_{10}^5 \times C_n^5$

   $$p_n = \frac{C_{10}^5 \times C_n^5}{C_{n+10}^{10}}.$$

   Si on étend les combinaisons, cela donne :

   $$p_n = \frac{(10!)^2}{(5!)^3} \times \frac{n(n-1)(n-2)(n-3)(n-4)}{(n+10)(n+9)(n+8)\dots(n+3)(n+2)(n+1)}.$$

2. *Étudier le sens de variation de la suite p_n et calculer $\lim_{n\to\infty} p_n$.*
   On étudie $\frac{p_{n+1}}{p_n} = \frac{(n+1)(n+1)}{(n+11)(n-4)} = \frac{n^2+2n+1}{n^2+7n-44} = \frac{n^2+7n-44}{n^2+7n-44} + \frac{-5n+45}{n^2+7n-44} = 1 + \frac{5}{n^2+7n-44}(9-n)$.
   Le terme $\frac{5}{n^2+7n-44}$ est positif (on a $n \geq 5$), donc on a croissance pour $n < 9$ et décroissance pour $n > 9$.

   La suite p_n est asymptotiquement décroissante et à termes positifs, elle admet donc une limite en $+\infty$. le numérateur est asymptotiquement équivalent à n^5, tandis que le dénominateur est asymptotiquement équivalent à n^{10}, donc p_n est asymptotiquement équivalent à $\frac{1}{n^5}$, qui tend vers 0. La limite de p_n en $+\infty$ est donc 0.

2 Loi binomiale

Exercice 3
On jette une fois un dé non pipé.

1. *Quelle est la probabilité d'obtenir la face 1 ?*
   $P("face1") = \frac{1}{6}$. Notons $p = \frac{1}{6}$

   Quelle est la loi de cet évènement ?
   Il s'agit d'une loi de Bernoulli de paramètre p, $\mathcal{B}e(p)$.

   On jette 18 fois le dé en question. Quelles sont les probabilités des évènements suivants :

2. *Obtenir exactement 3 fois la face 1*
   Si on connaît les lancers où se produisent l'événement "face 1", on a la probabilité p pour chacun de ces 3 lancers, et la probabilité $1-p$ pour chacun des $18-3=15$ autres. D'autre part, on a C_{18}^3 façons de choisir 3 lancers parmi 18, donc la probabilité d'obtenir exactement 3 lancers "face 1" est $C_{18}^3 \frac{1}{6}^3 \frac{5}{6}^{15}$.

   De manière générale, pour un événement suivant une loi de Bernoulli $\mathcal{B}e(p)$, le nombre de réussites X de cet événement lorsque l'on a n tentatives suit ce qu'on appelle une loi binomiale, avec

$$P(X=k) = C_n^k p^k (1-p)^{n-k}.$$

3. *Obtenir au moins 3 fois la face 1*
   $P(X \geq 3) = 1 - P(X=0) - P(X=1) - P(X=2)$ puis on applique ce qui précède.
4. *Obtenir au plus 16 fois la face 1*
   $P(X \leq 16) = 1 - P(X=17) - P(X=18)$ puis on applique ce qui précède.

3 La loi normale (gaussienne)

On suppose que $U \sim \mathcal{N}(0,1)$ (Loi normale centrée réduite).

Exercice 4 *Calculer les probabilités suivantes : $P(U < 1.5)$; $P(U > 2.5)$; $P(U < -1.5)$; $P(-1.5 < U < 2.5)$*

— $P(U < 1.5) = 0.9332$, par une lecture directe de la table.
— $P(U > 2.5) = 1 - P(U < 2.5) = 1 - 0.9938 = 0.0062$
— $P(U < -1.5) = P(U > 1.5) = 1 - P(U < 1.5) = 1 - 0.9332 = 0.0668$ par symétrie de la loi normale centrée réduite puis complémentarité
— $P(-1.5 < U < 2.5) = P(U < 2.5) - P(U < -1.5) = 0.9938 - 0.0668 = 0.9270$

Exercice 5 *Trouver la valeur de u telle que : $P(U < u) = 0.95$; $P(U < u) = 0.1$; $P(U > u) = 0.99$; $P(-u < U < u) = 0.95$*

— $P(U < u) = 0.95$: $u = 1.645$ par une lecture directe de la table (on prend la moyenne entre les deux valeurs pour lesquelles les probabilités encadrent 0.95)
— $P(U < u) = 0.1$: u est négatif, on se ramène alors au cas symétrique, $P(U < u) = P(U > -u) = 1 - P(U < -u)$. On cherche alors t tel que $P(U < t) = 1 - 0.1 = 0.9$, et $u = -t = -1.285$
— $P(U > u) = 0.99$: u est négatif encore une fois. $P(U > u) = P(U < -u) = 0.99$, on a donc $u = -2.325$
— $P(-u < U < u) = 1 - 2P(U > u) = 1 - 2(1 - P(U < u)) = 2P(U < u) - 1$. Remarque : $u > 0$ donc $P(U < u) > 0.5$. On cherche alors le u tel que $P(U < u) = \frac{1+0.95}{2} = 0.975$, d'où $u = 1.96$

On suppose que $X \sim \mathcal{N}(\mu = 2, \sigma^2 = 5^2)$ (Loi normale $\mathcal{N}(\mu, \sigma^2)$). On a alors $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Exercice 6 *Calculer les probabilités suivantes :* $P(X < 10)$ *;* $P(0 < X < 10)$

On pose $U = \frac{X-\mu}{\sigma} = \frac{X-2}{5}$. On va se ramener à des calculs sur U, puis en déduire ce qu'il faut sur X.

— $P(X < 10) = P(X - 2 < 8) = P\left(\frac{X-2}{5} < 1.6\right) = P(U < 1.6) = 0.9452$ par une lecture directe de la table.

— $P(0 < X < 10) = P(-0.4 < U < 1.6) = P(U < 1.6) - P(U < -0.4) = P(U < 1.6) - (1 - P(U < 0.4)) = P(U < 1.6) + P(U < 0.4) - 1 = 0.9452 + 0.6554 - 1 = 0.6006$

Exercice 7 *Trouver la valeur de* x *telle que :* $P(X < x) = 0.95$ *;* $P(X < x) = 0.05$ *;* $P(2 - x < X < 2 + x) = 0.95$

— $P(X < x) = P(U < \frac{x-2}{5}) = 0.95$, donc $\frac{x-2}{5} = 1.645$, c'est-à-dire $u = 5 \times 1.645 + 2 = 10.225$.

— $P(X < x) = P(U < \frac{x-2}{5}) = 0.05$: on pose $-t = \frac{x-2}{5}$. $P(U < -t) = 1 - P(U < t)$. On cherche donc t tel que $P(U < t) = 1 - 0.05 = 1.95$, donc $t = 1.645$. Donc $\frac{x-2}{5} = -1.645$, c'est-à-dire $u = -5 \times 1.645 + 2 = -6.225$.

— $P(2 - x < X < 2 + x) = P(-t < U < t) = 0.95$ avec $t = \frac{x}{5}$. On a déjà calculé cela dans l'exercice précédent : $t = 1.96$, donc $u = 5 \times 1.96 = 9.80$.

4 La loi du Chi-deux χ^2 (ou loi de Pearson)

$U_1, \ldots, U_p$ étant p variables $\mathcal{N}(0, 1)$ indépendantes, on appelle loi du chi-deux à p degrés de liberté (χ^2_p) la loi de la variable $\sum_{i=1}^{p} U_i^2$.

Exercice 8

1. *Que vaut la somme de deux variables de* χ^2 *indépendantes à* p *et* q *degrés de liberté.*
   On considère que X suit une loi du χ^2 à p degrés de libertés, et Y suit une loi du χ^2 à q degrés de libertés, que X et Y sont indépendantes. X suit la loi de $\displaystyle\sum_{i=1}^{p} U_i$, où les U_i sont des variables aléatoires indépendantes suivant une loi normale centrée réduite. De même Y suit la loi de $\displaystyle\sum_{i=1}^{q} V_i$, où les V_i sont des variables aléatoires indépendantes entre elles et indépendantes de U_i, suivant une loi normale centrée réduite. Alors $X + Y$ suit la loi de la somme de $p + q$ variables aléatoires indépendantes suivant une loi normale centrée réduite, donc une loi du χ^2 à $p + q$ degrés de liberté.
2. *Soit* $X \sim \chi^2_{15}$ *et* $Y \sim \chi^2_{10}$*. Calculer* $P(X < 6.26)$*,* $P(Y > 3.25)$*,* $P(X + Y > 11.52)$. cf. Tables ou TP.
3. *Soit* $X \sim \chi^2_{15}$*. Trouver* x *tel que* $P(X < x) = 0.01$*,* $P(X < x) = 0.05$*,* $P(X < x) = 0.99$. cf. Tables ou TP.

5 Théorème de la limite centrale

Quelques propriétés de la loi normale

1. Linéarité
   Soit $X \sim \mathcal{N}(\mu, \sigma^2)$. La transformation linéaire $Y = aX + b$ où a et b sont deux réels, définit une variable aléatoire Y de loi normale $\mathcal{N}(a\mu + b, a^2\sigma^2)$.
2. Somme
   La somme de deux variables aléatoires normales est normale. La moyenne de la somme est la somme des moyennes. La variance de la somme est la somme des variances si les variables sont indépendantes (dans le cas contraire cela est un peu plus compliqué).

Exercice 9 *Soit X de loi $\mathcal{N}(m, \sigma^2)$. On pose $Y = aX + b$ où a et b sont deux réels.*

1. *Que vaut $E(Y)$?*
2. *Que vaut $V(Y)$?*

cf. Paragraphe sur la linéarité.

Exercice 10 *Soit $X_1, \ldots, X_n$ n variables aléatoires normales $\mathcal{N}(m, \sigma^2)$ et indépendantes. On note $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ la moyenne empirique. Quelle est la loi de $\bar{X}$?*

En généralisant par récurrence la propriété sur la somme, $\sum_{i=1}^{n} X_i$ suit une loi normale $\mathcal{N}(nm, n\sigma^2)$. Par linéarité, $\bar{X}$ suit une loi normale $\mathcal{N}(m, \frac{\sigma^2}{n})$

Le théorème de la limite centrale généralise le résultat précédent à une variable aléatoire X quelconque. Si $X_1, \ldots, X_n$ est une suite de variables aléatoires de même loi et indépendantes, ayant une moyenne théorique m et un écart-type σ, alors la moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ suit approximativement la loi normale $\mathcal{N}(m, \frac{\sigma^2}{n})$.

Exercice 11 *Appliquer ce théorème au cas de n variables indépendantes de Bernoulli $\mathcal{B}e(p)$. Que vaut $\bar{X}$? En déduire une approximation de la loi binomiale par une loi normale.*

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ où les X_i suivent une loi $\mathcal{B}e(p)$. D'après le théorème de la limite centrale, $\bar{X}$ suit approximativement la loi normale $\mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$. La loi binomiale à n tirages peut alors être approximée par l'approximation de $n\bar{X}$, c'est-à-dire une loi normale $\mathcal{N}\left(np, np(1-p)\right)$.

Exercice 12 *(Application de l'exercice précédent) Mathieu lance 100 fois une pièce de monnaie et compte le nombre de fois où il obtient "face".*

Le nombre de « face » X obtenu en 100 lancers suit une loi binomiale $\mathcal{B}i(n = 100; p = 0,5)$.Nous sommes dans les conditions d'approximation de la binomiale par une loi normale de moyenne $m = np = 50$ et d'écart type $\sigma^2 = npq$.

— *Quelles sont les probabilités :*

1. *Qu'il obtienne moins de trente fois "face" ?*
   Cela revient à calculer $P(X \leq 30)$ (ou $P(X < 30)$). Pour faire le calcul, on utilise l'approximation par une loi normale que l'on réduit :
   $$X \leq 30 \Leftrightarrow \frac{x-50}{5} \leq \frac{30-50}{5} \Leftrightarrow U \leq -4,$$
   où U suit une loi normale centrée réduite. Une lecture de table donne $P(U \leq -4) = 32.10^{-6}$.

   Note : Calcul avec correction.
   Si l'on veut opérer une correction, il convient dans ce cas de distinguer les deux événements $X < 30$ et $X \leq 30$. En utilisant l'approximation par une loi normale, on remplace $X < 30$ par $X < 29,5$ et $X \leq 30$ par $X \leq 30,5$. On obtient alors : $P(X < 29,5) = P(U < -4.1) = 26.10^{-6}$ (4,1 n'est pas dans la table : on a procédé à une interpolation entre 4 et 4,5). Et $P(X < 30,5) = P(U < -3,9) = 72.10^{-6}$.
2. *Qu'il obtienne un nombre de "face" compris entre 40 et 60 ?*
   Calculons $P(40 \leq X \leq 60)$. En approximant X et en réduisant on obtient : $P(40 \leq X \leq 60) = P(-2 \leq U \leq 2) = 95,44\%$.

— *Trouver le plus petit nombre entier n tel que Mathieu ait au plus une chance sur 100 d'avoir un nombre de "face" supérieur ou égal à n.*
L'entier n cherché doit vérifier $P(X \geq n) \leq 1\%$. En approximant et en réduisant on obtient :
$$P(X \geq n) \leq 1\% \Leftrightarrow P\left(U \geq \frac{n-50}{5}\right) \leq 1\%.$$
Une lecture de table indique alors qu'une condition nécessaire et suffisante pour que l'inégalité ci-dessus soit satisfaite est que $\frac{n-50}{5} \geq 2,3263$ ou encore, n étant entier que $n \geq 62$. Le plus petit entier est donc $n = 62$.

— *Mathieu, après avoir lancé 100 fois sa pièce, a obtenu 66 fois "face". Peut-on le soupçonner d'avoir triché ?*
En toute rigueur, on ne peut pas dire que Mathieu a triché, mais au vu du résultat précédent on sait que Mathieu a au plus une chance sur 100 d'obtenir un résultat supérieur à 62. A ce titre, mais on prend le risque de se tromper, on peut le soupçonner d'avoir tricher (le risque pris peut s'apprécier puisque dans moins de 1 cas sur 100 nous saurons avec une telle configuration si Mathieu n'a pas triché ! Cette démarche sera explicitée avec la notion de test).

6 Un petit tour en Suisse

On suppose que les hommes suisses de plus de 50 ans ont des poids distribués approximativement selon une loi normale avec une moyenne théorique $\mu = 70kg$ et un écart-type $\sigma = 5kg$.

Exercice 13 *1.* Si on tire au hasard un individu dans cette population, calculer la probabilité pour qu'il ait :

(a) un poids compris entre 75 et 80kg.
On considère X la variable aléatoire de loi normale $\mathcal{N}(70, 25)$.

$$P(75 < X < 80) = P\left(\frac{75-70}{5} < Z < \frac{80-70}{5}\right) \text{ où } Z \rightsquigarrow \mathcal{N}(0,1)$$

Soit : $P(75 < X < 80) = P(1 < Z < 2) = P(Z < 2) - P(Z < 1) = 0.1359$

(b) un poids excédant 85kg.
Par le même genre de transformation, et une symétrie : $P(X > 85) = P(Z > 3) = 0.00135$

2. Quel poids dépassent les 10% des individus les plus lourds de cette population ?
On cherche ici le quantile d'ordre 0.1 de la loi :

$$P(X > t_{0.1}) = P\left(Z > \frac{t_{0.1} - 70}{5}\right) = 0.1$$

Or $u_{0.1} = 1.281552$ *(obtenu dans les tables par symétrie...). Donc* $t_{0.1} = 76.40776$.

3. On extrait au hasard un échantillon de 25 individus. Quelle est la probabilité pour que la moyenne de cet échantillon dépasse 85kg ?
On considère la statistique $\overline{X} = \sum_{i=1}^{25} X_i$ *où les* X_i *sont des v.a.i.i.d. de loi* $\mathcal{N}(70, 25)$. *La loi de* $\overline{X}$ *est alors* $\mathcal{N}(70, \frac{25}{25} = 1)$. *Donc* $P(\overline{X} > 85) = P\left(Z > \frac{85-70}{1}\right) = P(Z > 5) = 2.866516e - 07$.