Séance 1 : Introduction aux lois de probabilité et utilisation d'Excel

Connaissances préalables : notion de moyenne

Matériel : une calculatrice scientifique et un ordinateur disposant d'Excel ou d'un tableur équivalent

Une brève introduction

Toute approche déterministe rencontre une limite fondamentale : pour avoir un modèle complet de la réalité permettant de faire la prévision ou de l'analyse, il faudrait prendre en compte un nombre considérable de facteurs. Par exemple, pour savoir combien de personnes vont être présentes dans la file d'attente d'un bureau de poste pour un jour donné, il faut savoir pourquoi chaque personne vient, connaître l'état de sa motivation, le retard éventuel qu'elle peut subir dans les transports, le fait que certaines personnes vont remettre leur sortie à la Poste s'ils constatent qu'il pleut... Il est totalement impossible de construire un modèle déterminste avec autant de facteurs. De fait, la réalité est souvent approchée au travers de modèles probabilistes.

Pour certains phénomènes, on sait qu'ils peuvent être modélisés par telle ou telle Loi de probabilité. Par exemple, pour les phénomènes de file d'attente, on dispose de la Loi de Poisson. Supposons que l'on gère un central d'appel, à partir du nombre moyen de personnes appelant le central que l'on constate, on pourra établir la probabilité que le standard soit saturé, la probabilité qu'il soit en sur-capacité et donc déterminer le nombre nécessaire des répondants.

Pour certains phénomènes, on ne dispose pas de loi de probabilité déterminée. On peut établir la moyenne et l'écart type de ces phénomènes et dans ce cas, on pourra, sous certaines conditions, les représenter par une Loi Normale.

Notions préalables

Factorielle

La notation $n!$ se lit "factorielle n". On la définit de la façon suivante : $n! = n*(n-1)*(n-2)*(n-3)*...*2*1$. Par convention, $0! = 1$ (en réalité, il ne s'agit pas tout à fait d'une convention, mais on l'admettra).

1- Donner la valeur de $3!$, de $6!$, de $8!$ en les calculant manuellement

2- En utilisant Excel ou votre calculatrice, calculer la valeur de $65!$, jusqu'à quel nombre peut on calculer la factorielle avec l'un et l'autre de ces deux outils ?

Combinaison (à vérifier)

La notation C_n^p correspond à $\frac{n!}{!p*(n-p)!}$. Cette formule correpond à la manière de choisir p éléments parmi n. Par exemple, supposons un groupe de 20 étudiants, on veut constituer un groupe de 5 personnes : on a C_{20}^5 groupes différents qu'il est possible de constituer. Combien de paires d'étudiants pourrait-on constituer à partir du même groupe ? $C_{20}^2=\frac{20!}{2!*18!}=\frac{20*19}{2}=190$ (à noter le fonctionnement de la simplification).

3- Compter le nombre de paires que l'on peut constituer au sein d'un groupe de 4 personnes, d'abord sans utiliser la formule du nombre de combinaisons, puis en utilisant cette formule.

4- Soit un groupe de n=30 personnes, combien y a t'il de combinaisons de 5 personnes ? En utilisant Excel, déterminer le nombre des combinaisons possibles pour toutes les valeurs de p=1 à p=30 et tracer le graphique qui montre l'évolution du nombre des combinaisons en fonction de p. Pour quelle valeur de p le nombre de combinaison est il maximal ? Comment pourrait-on l'expliquer ?

5- [A passer, sauf goût particulier pour les mathématiques] Puisque la formule qui donne le nombre de combinaisons contient une division, peut on avoir un nombre de combinaisons qui ne serait pas un entier ?

Variance et écart type

On considère deux séries de données :

9 10 11 et 5 10 15.

6- Donner les moyennes de chacune de ces séries de données.

Malgré leur moyenne commune, quelle constat doit on faire entre ces deux séries de données ?

7- Quelle donnée statistique permet de prendre en compte la différence entre ces deux données ?

8- Calculer la variance et l'écart type de ces deux séries de données ?

9- Calculer la variance à la main. Comment pourrait on la définir ?

10- Considérez une série de données assez conséquentes et assurez vous de savoir calculer moyenne et écart type avec votre calculatrice.

11- A partir de la définition de la variance, déterminer quelle est la fonction qui permet de calculer la variance exacte entre VAR et VAR.P sur Excel.

12- Faire un tirage de données aléatoires sur Excel en utilisant la fonction ALEA() : quelle moyenne et quel écart-type constate t-on ?

Lois de probabilité discrètes de référence

Quand on ne pas connaître de manière déterministe une donnée (le nombre de personnes qui viennent à un guichet, le nombre de produits défectueux sur une série, le débit d'une rivière...), on la représentera comme une variable aléatoire.

A défaut de connaître sa valeur ou ses valeurs exactes, on cherche avec quelle probabilité cette variable prend telle ou telle valeur (avec quelle probabilité a t'on plus de 100 personnnes qui appelent le standard, avec quelle probabilité plus de 1000 produits seront défectueux etc...). Ici, on parle de variables discrètes : c'est à dire que ces variables peuvent seulement prendre un nombre limité de valeurs (en terme formel, on dira qu'elles peuvent prendre un nombre dénombrable de valeurs).

Pour chaque loi, on explique son schéma : ie le cas dans lequel elle s'applique. Si certaines conditions sont respectées, on peut dire que telle Loi s'applique.

Loi binomiale

Dans le cadre de la loi Binomiale, on a une série de n expériences indépendantes. Pour chaque expérience, elle a deux issues possibles (succès/échec, sain/malade, blanc / noir) . Une variable qui suit une loi de Binomiale va être par exemple le nombre de succès ou d'échec sur les n expériences.

Exemple de situations :

- On considère un groupe de 100 patients, on sait que chacun a une probabilité p d'avoir contracté une maladie. On veut savoir quelle est la probabilité que 6 patients aient cette maladie, combien de partients ont contracté en moyenne la maladie etc...
- On considère 1000 produits fabriqués à la chaîne, on sait qu'il a une probabilité p=0,03 pour chaque produit d'être déecteux. Combien de produits seront alors défectueux en moyenne, quelle est la probabilité d'avoir plus de 10 produits défectueux etc...

A noter que les termes succès ou échec sont utilisés ici sans à comprendre sans aucune connotation ici.

Soient n expériences i**ndépendantes**, chaque expérience ayant une probabilité p de succès (et donc une probabilité 1-p d'échec), la variable aléatoire X est le nombre de succès, la probabilité d'avoir k succès est donnée comme :

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

La moyenne de X est $n*p$

La variance de X est $n*p*(1-p)$

13- Dans le cadre d'une pharmacie, on veut prévoir le nombre de boîtes de médicament C qu'il faut stocker pour la journée (le réapprovisionnement de ce produit étant quotidien). On peut tabler sur la présence de 70 clients chaque jours. Il y a une probabilité p=0,03 qu'un client se soit fait prescrire le médicament C.

- Définir la notion de succès ou d'échec : expliquer pourquoi on peut utiliser la Loi Binomiale ici
- Quel est le nombre moyen des clients qui vont venir avec une prescription de médicament C sur la journée ?
- Quelle est la probabilité que 8 patients viennent avec une prescription de médicament C ? la probabilité de 10 partients ? La probabilité de 70 patients ? La probabilité de 0 patients ?
- Quelle est la probabilité que moins de 8 patients viennent avec une prescription de médicament C (on pourra utiliser Excel) ?
- Quelle est la probabilité que plus de 30 patients viennent avec un prescription de médicament C (on pourra utiliser Excel) ?
- Sachant que la pharmacienne considère que si le risque de manquer de boites représente seulement une probabilité de 0,01, c'est un risque acceptable, combien de boîtes de médicaments devrait elle stocker (on pourra utiliser Excel) ?

Loi de Poisson

La Loi de Poisson modélise "le comportement du nombre d'évènements se produisant dans un laps de temps fixé" (Wikipedia). On va considérer une variable X qui est le nombre de fois où un évenement va se produire dans un laps de temps donné, **sachant que l'on sait quelle est le nombre moyen de fois où cet évenement se produit sur ce laps de temps**. On nombre λ c e nombe moyen.

Dans ce cas, on sait que l'on a :

$$p(X=k)=e^{-\lambda}*\frac{\lambda^k}{k!}$$

La moyenne de X est λ

La variance de X est λ

14- Au guichet d'un service public, on sait que 7 personnes se présentent chaque 10 minutes, en moyenne. On veut être quasiment sûr qu'une personne donnée n'aura pas à attendre lorsqu'elle arrive.

- Expliquer pourquoi une approche déterministe est insuffisante. Quel nombre de guichetiers recommanderaiton en l'absence d'information sur la Loi statistique ?

- Expliquer pourquoi on retrouve le schéma de la Loi de Poisson et quelle est la valeur du paramètre de cette Loi.
- Donner les probabilités repesctives que 0, 5, 7, 9, 14 personnes se présentent au guichet en 10 minutes
- Programmer, sur un tableur, la loi de probabilité, ie faire un tableau avec pour chaque valeur du nombre k de personnes se présentant, la probabilité que cet évenement arrive. Faire que le paramètre de la fonction soit saisi dans une cas à part et puisse être modifié.
- On veut être assuré, avec une probabilité supérieure à 0,95, que tout nouveau usager qui se présentera n'attendra pas : combien faut-il avoir de guichets ouverts ?
- Est ce que les conclusions seraient changées si le nombre moyen de personnes se présentant chaque 10 minutes était 8 ?
- Quelle est la probabilité que moins de 5 personnes se présentent dans dans une tranche de 10 minutes ?

A noter qu'en donnant pour chaque valeur possible de la variable, la probabilité qui lui est associée, on donne la **distribution** de la variable.

Notion de distribution de probabilité discrète

Une variable aléatoire est dite discrète lorsque elle prend un nombre fini ou dénombrable de valeurs. Sans rentrer dans les détails, une variable aléatoire qui correspond aux notes au sein d'une classe, ou un nombre de personnes présentes à un guichet, ou un nombre de produits défectueux sont des variables aléatoires discrètes.

Distribution de probabilité

Etablir la distribution de probabilité d'une variable aléatoire discrète consiste à déterminer pour chaque valeur possible de la variable, la probabilité qui lui est associée. On peut représenter la distribution de probabilité avec en abscisse les valeurs possibles de la variable et en ordonnées la probabilité.

14- En utilisant un tableur, faire apparaître la distribution de probabilité d'une Loi Binomiale et d'une loi de Poisson. Faire en sorte que les paramètres des fonctions soient placés dans des cases et puissent être modifiés afin de voir l'évolution de la représentation graphique avec l'évolution de la valeur de ces paramètres. Utiliser une feuille de calcul par loi de probabilité.

Fonction de répartition

La fonction de répartition est la fonction qui donne en fonction de chaque valeur k de la variable aléatoire, la probabilité que la valeur de la variable aléatoire soit inférieure : $p(X \leq k)$

15- Pour chacune des lois de probabilités, dans la feuille de calcul qui lui est consacrée, donner la représentation de la fonction de répartition.

Loi de probabilité empirique

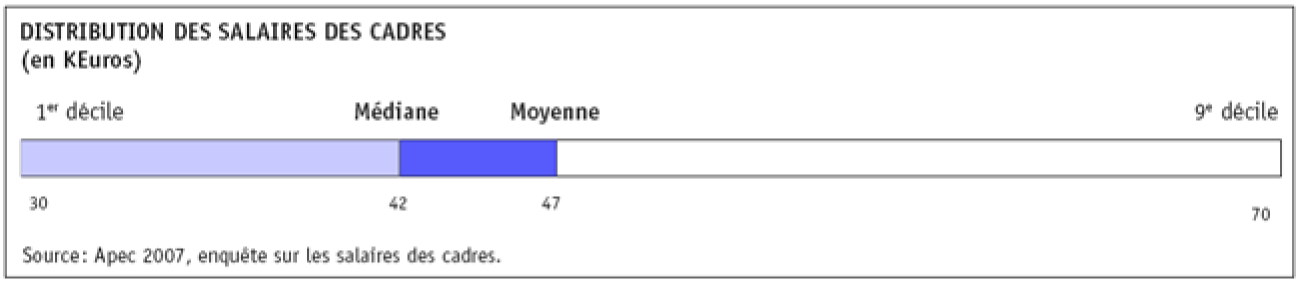
Au vu de la description des phénomènes précédents, on comprend que certains phénomènes sont identifiés comme correspondant à une loi de probabilité précise. Pour beaucoup d'autres phénomènes, ils n'ont pas nécessairement une loi de probabilité déterminée ou bien ils évoluent en fonction de lois qui ne sont pas celles données dans la partie précédente.

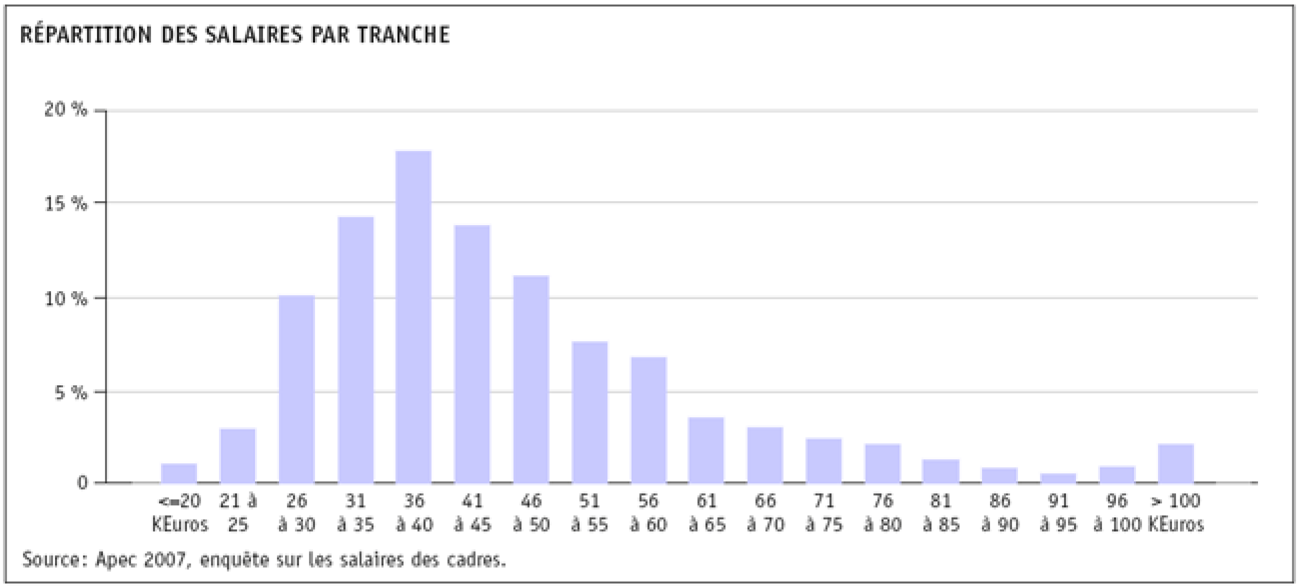
Il est bien sûr possible de construire des distributions de probabilités empiriques : ie des distributions de probabilité à partir des données dont on dispose sur un phénomène et non pas des distributions théoriques, dépendantes d'une loi. On peut très bien partir de la série des notes dans une classe, de la série des températures à Paris sur l'année, de la série des niveaux d'enneigements au cours des 10 dernières années et tenter d'en constituer une distribution de fréquences, ie de probabilité

Lois de probabilités continues et distribution de probabilité

A la notion de variable aléatoire discrète s'oppose la notion de variable aléatoire continue, qui peut donc prendre un nombre non dénombrable de valeurs : la température, l'hydrométrie, le poids des indivdus etc... De la même manière, on a des lois qui définissent des phénomènes dans le cas des variables aléatoires continues (cf la présentation de la loi normale ci dessous) et on peut établir, pour les données qui correspondent à des variables aléatoires continues, des distributions, fonction de répartition etc...

Par exemple :





A noter qu'ici, on donne non pas des valeurs en abscisses, mais des catégories.

16- A partir d'une série statistique quelconque librement récupérée sur internet (PIB, série de température etc..) et comportant au moins 100 valeurs, établir la distribution des données correspondantes : sous forme d'histogramme et sous forme de courbe.