Notions de Probabilités.

Préambule.

L'étude des probabilités, autrefois réservée aux quelques professions qui en avaient besoin, est maintenant au programme, dès le lycée.
A lire les différents exercices, il apparaît que cette spécialité mathématique est surtout utilisée comme support intellectuel, probablement au détriment d'autres spécialités, comme la géométrie.

Le but de ce papier est de présenter les notions de base fondamentales qui sont celles d'une science fondée sur le réel.
Volontairement, je n'ai indiqué aucune référence. Il en existe beaucoup, j'ai mes préférences, et j'y ferai allusion lorsque l'occasion se présentera.

Sauf cas particulier, les notions de mathématiques élémentaires sont supposées connues ainsi que les formules d'analyse combinatoire : arrangements, permutations, combinaisons.

Je m'attacherai à donner une définition précise des termes que j'emploie, sachant que cette définition n'est pas la seule qui puisse être donnée. Ce papier ne doit en aucun cas être considéré comme un cours.
Les phrases et paragraphes notés entre guillemets sont issues du cours de J.J. LEVALLOIS (1960).

Probabilité : un nombre compris entre 0 et 1 qui est le rapport du nombre de cas favorables sur le nombre de cas possibles.
Ex.: "Tirer un As dans un jeu de 52 cartes".
Nombre de cas favorables = 4.
Nombre de cas possibles = 52.
Probabilité = 4/52 = 1/13 ~ 0.077.
A noter que la valeur d'une probabilité est un nombre exact, parfaitement connu dans le cas présent (1/13). La valeur 0.077 est une valeur approchée, le signe '~' signifie "approximativement égal".
Une probabilité égale à 0 correspond à un événement impossible. Une probabilité égale à 1 correspond à un événement certain.

Il est toujours possible de me contacter pour faire une observation, poser une question, proposer un exemple. Les cas que je jugerai intéressants seront cités à la fin du chapitre concerné.

Postulat de la moyenne

Mesure : C'est la valeur observée d'une quantité mesurable. On appelle quantité mesurable une quantité à laquelle on peut appliquer les opérations arithmétiques, par exemple, une longueur, un volume, une durée, un nombre d'éléments. Une mesure de température est dite non mesurable, par exemple 20° + 10° ne font pas 30°.

Observation : C'est un terme plus général que mesure, et dans le contexte des probabilités, il peut être considéré comme synonyme. On pourra trouver des cas où cette distinction peut être nécessaire. Par exemple, on observe un certain phénomène mesuré par un comptage de nombre d'éléments, mais intervient aussi un facteur climatique que l'on cote de 1 à 5, 1 : mauvais temps ; 5 beau temps.

Moyenne : En mathématiques, il existe plusieurs moyennes (pondérée, géométrique, arithmétique, harmonique, quadratique). En l'absence de qualificatif, il s'agit de la moyenne arithmétique, égale à la somme des termes divisée par leur nombre.

Lorsqu'on effectue plusieurs mesures d'une même quantité, en considérant que ces mesures ont été faites avec le même soin et les mêmes méthodes, on peut se demander quelle est la valeur à adopter. L'intuition nous a fait choisir la moyenne arithmétique. On aurait pu choisir une autre moyenne, par exemple la moyenne géométrique, ou une autre valeur qui ne serait pas forcément une moyenne.

Supposons un grand nombre de mesures d'une même choses, faites dans les mêmes conditions. On classe les résultats par ordre croissant par exemple. Nous constatons que toutes les valeurs sont dispersées entre les deux extrêmes et qu'entre ces deux bornes la répartition des valeurs n'est pas uniforme : la "densité" de ces mesures va en croissant du terme A jusque vers le milieu de l'intervalle complet AB puis décroît en passant sensiblement par les mêmes valeurs.
Les points représentatifs admettent à première vue un point d'accumulation vers le milieu de l'intervalle de répartition.
Si l'on groupe les mesures par couple à partir des extrémités,
M(1) + M(n)
M(2) + M(n-1)
……
M(p) + M(n-p)
On constate que ces sommes sont à peu près égales et que (M(p) + M(n-p))/2 a sensiblement la valeur trouvée au point où l'accumulation des mesures est la plus grande.

On appelle "Postulat de la moyenne" la certitude que la moyenne arithmétique des mesures observées est une valeur très proche de la mesure recherchée, étant donné l'ensemble des éléments dont on dispose. Le terme "postulat" ne doit pas être pris à la légère. Au stade où on en est, on ne sait pas le démontrer, mais un grand nombre d'observations et de vérifications justifient ce choix. On verra plus loin que l'on peut démontrer qu'on a fait le bon choix.

Le hasard

C'est une notion difficile et il est nécessaire de s'y attarder.
Le Larousse dit : "Cause attribuée aux évènements considérés comme inexplicables logiquement et soumis seulement à la loi des probabilités".
John Hartong, dans son livre "Probabilités et Statistiques" explique en détail et très clairement ce qu'est le hasard et je ne vais pas le paraphraser.

Le hasard a quelque fois été comparé, voire confondu, à la notion de chaos, c'est une erreur fondamentale, nous verrons plus loin pourquoi.

En supposant qu'aucun élément connu n'intervienne dans la chute d'un astéroïde, si un astéroïde tombe dans mon jardin, on dira que c'est le hasard.

John Hartong utilise le "Paradoxe de Bertrand" dans son introduction et je vais résumer le dit-chapitre.
La question posée est : "Soit un cercle et une corde de ce cercle. Quelle est la probabilité que la longueur de la corde soit supérieure au côté du triangle équilatéral inscrit ?". On connaît trois méthodes de calcul possibles, les résultats sont ½, 1/3 et ¼. On pourrait d'ailleurs en imaginer d'autres.
Cela voudrait-il dire que le problème tel qu'il est posé n'a pas de solution ? Il manquerait donc une information ? Cette information serait la réponse à une question du type "quel hasard ?" ou exprimé plus mathématiquement "quelle loi de probabilité doit-on utiliser pour savoir de quel hasard il s'agit ?". En d'autres termes il y aurait plusieurs hasards, un nombre fini ou pas ?
Tout cela n'est intellectuellement pas très sérieux.

Le présent paragraphe ne vaut en aucun cas démonstration. Ce n'est qu'une approche de la notion de hasard.

Distribution uniforme

Cette expression n'a pas encore été employée, mais elle est sous-entendue dans tout le texte, et il est temps de la préciser.

Expérience : J'appelle expérience, quelle que soit la nature et quel que soit le contexte, la réalisation d'une suite d'opérations élémentaires, indépendantes,

effectuées dans les mêmes conditions ou dans des conditions strictement équivalentes. Une expérience très simple et à laquelle on se réfère souvent est le jeu à pile ou face. L'opération élémentaire est un jet de pièce, l'expérience est un ensemble de jets de pièces.

Lorsqu'on fait une expérience, celle-ci n'est valable que si les différents éléments mesurés ou observés sont indépendant et dans un contexte identique. On appelle cela généralement "variables aléatoires indépendantes identiquement distribuées". Cela peut se concrétiser par le fait que tous les éléments mesurés ou observés sont choisis dans un ordre quelconque. Il sont tous interchangeables et seul le résultat de la mesure ou de l'observation est à prendre en considération.
A l'inverse, si on constate une évolution systématique en rapport direct avec l'ordre des observations, alors on n'est plus dans le contexte de l'expérience du présent papier.

Il peut arriver que les différents éléments d'une expérience soient comparable à une nuance près qui pourrait être que ceux-ci sont réalisés dans des conditions légèrement différentes. On tourne la difficulté en appliquant un poids aux résultats. On utilisera, dans ce cas, une moyenne pondérée. Le but de cette opération est de respecter la condition indispensable que j'ai appelée "contexte identique". Voir le paragraphe des modèles pour plus de détails.

Dispersion des écarts à la moyenne.

On est donc en présence d'une série d'observations d'événements indépendants et identiquement distribués. Puisque ces éléments sont indépendants, leur apparition est aléatoires et régie par le hasard. A chaque élément correspond une valeur qui est sa mesure.
Conformément au postulat de la moyenne, on calcule la moyenne arithmétique de ces mesures.
Pour chaque mesure, calculons l'écart à la moyenne. La somme algébrique de ces écarts est naturellement 0. Divisons l'intervalle [écart max – écart min], par exemple, par 10 et classons ces écarts dans chaque classe ainsi déterminée. Enfin, reportons sur un graphique, en abscisse les limites de classe et en ordonnée le rectangle dont la hauteur sera le nombre d'écarts dans la classe correspondante.
Si l'on essaye de tracer une courbe joignant par un trait continu les points moyens des rectangles ou plus exactement laissant de part et d'autre des aires égales, on obtient une courbe ayant la forme d'une cloche.

Cette expérience peut être faite dans n'importe quel contexte et la courbe observée aura toujours le même forme.

Prenons l'exemple du jeu de pile ou face. La fonction random() permet de réaliser cela facilement. Voir le paragraphe sur la fonction random() pour plus de détails.
A partir d'une même simulation, on va faire deux expériences.

1- on compte le nombre de pile et le nombre de face,
2- on groupe les résultats successifs par 4, on obtient ainsi un nombre binaire, si on attribue 0 à pile et 1 à face. Ce nombre peut valoir de 0 à 15 en base décimale.

Pour le cas 1, on vérifie rapidement que le nombre de pile est très voisin du nombre de face. Graphiquement, ce n'est pas très spectaculaire, mais il est indispensable de le vérifier et de le constater.
Pour le cas 2, on a 16 issues possibles. La moyenne étant naturellement le nombre de groupes de 4 divisé par 16.
Si on compte le nombre d'occurrences de chacune des issues et que l'on calcule le nombre d'écarts à la moyenne pour chaque occurrence, on peut établir des classes et dessiner le graphique tel qu'il a été décrit précédemment. On constatera que si le nombre de jet de pièce est suffisamment grand, la courbe représentative est telle qu'on l'attendait.

Fonction Random()

Le but de la fonction random est de générer des nombres aléatoires avec un ordinateur.
On a démontré que, strictement parlant, ce n'était pas possible pour les 2 raisons suivantes :

1- quelle que soit la méthode utilisée, au début de la liste, l'ordinateur donnera toujours le même nombre.
2- Au bout d'un certain temps, le cycle de sortie des nombres recommencera de la même façon. En d'autres termes un nombre N tiré au rang n sera toujours suivi du même nombre M au rang n+1.

Pour résoudre le point 1, on a introduit ce qu'on appelle la graine. C'est un nombre initialisé avec un élément extérieur, généralement l'horloge de la machine.
Concernant le point 2. La plupart des fonctions random ont un cycle de 2^32, ce qui vaut environ 4 milliards, c'est à dire largement suffisant tant qu'il ne s'agit pas de traitement sensible, par exemple jeu ou cryptographie. Pour les traitements dit sensibles, il existe des fonctions dont le cycle est environ 2^1024.

Dans tous les cas, ces fonctions sont dites "pseudo-aléatoires".

On peut trouver deux exceptions à cela.
La fonction GenRand(). Cette fonction génère à chaque tirage un nombre tel que la répartition de la loi normale (voir plus loin) sera strictement respectée.

La fonction que l'on peut trouver actuellement sur le Net n'est pas celle que j'avais testée. Je n'en dirai donc pas plus avant d'avoir fait des tests avec cette nouvelle version. A mon avis, cette fonction est à éviter.

Les fonctions de simulation de liste en matière de probabilité des logiciels orientés mathématique (Scilab, Matlab etc.).
Ces fonctions ne sont pas des générateurs de nombres pseudo-aléatoire sauf si le paramètre nécessaire est précisé, c'est à dire pas "par défaut". Ceci dépasse le cadre du présent papier.

Vérifications.

La première vérification que je présenterai est connue sous le nom de "Problème de l'aiguille" dû probablement à Buffon.

"On jette une aiguille sur une feuille horizontale de papier sur laquelle sont tracées des lignes parallèles équidistantes. Quelle est la probabilité pour que l'aiguille touche l'une des lignes ?"
Je ne détaillerai pas le calcul, je donnerai seulement la conclusion écrite par J.J. Levallois.
"[…]la probabilité cherchée est finalement $P=2l/\pi a$.
Elle dépend du nombre π… On a exécuté pratiquement cette expérience pour savoir si elle s'accordait avec la théorie. Le résultat constitue une éclatante vérification de celle-ci : sur un nombre de 5.000 épreuves on a calculé pour π la valeur 3,15. Il est difficile d'obtenir mieux dans la vérification des lois physiques : avec une aiguille de 36mm et un écart de lignes de 45mm, Wolf a obtenu : 2532 sécances au lieu de 2546 calculées : $1/\pi = 0.3169$ au lieu de 0.3183."

La seconde vérification est plus visuelle : la planche de Galton.

Sur une planche de bois, on plante des clous en ligne de façon que les clous de la ligne suivante soient décalés d'un demi intervalle par rapport à ceux de la ligne précédente.
La planche est inclinée de façon qu'une bille puisse la parcourir par la seule gravité. A partir d'un point situé plus haut de la première ligne, on laisse rouler une bille. Au premier clou rencontré, elle peut partir soit à droite soit à gauche. A la ligne immédiatement inférieure, elle rencontre un nouveau clou et peut donc dévier, soit à droite, soit à gauche, et ainsi de suite, jusqu'à la dernière ligne de clous, alors elle tombe dans un godet dont la largeur correspond au diamètre de la bille.
On répète l'opération avec un grand nombre de billes identiques. Elles s'accumulent les une au-dessus des autres dans les godets correspondant à la sortie de la dernière ligne. (On trouve de très jolies simulations sur le Net).

Si on observe de quelle façon se remplissent les godets, on remarque que la courbe qui joint la bille supérieure de chaque godet a, à tout moment, la forme d'une cloche, telle qu'on l'a observée lors du calcul de la répartition des écarts à la moyenne.

Pour mémoire, et sans détails supplémentaires, on observe cette courbe sur le seuil de vieilles maisons, usé par les nombreux passages.

Théorème de Bernoulli ou loi des grands nombres.

(Volontairement j'ai modifié l'orthographe du cours de J.J. Levallois).

"La fréquence d'un événement tend vers sa probabilité lorsque le nombre des épreuves devient très grand".

Je ne donnerai pas la démonstration, par contre, les possibilités offertes par les ordinateurs permettent de faire facilement une vérification par simulation.

Reprenons le jeu de pile ou face. On va étudier l'événement suivant : "suite continue de N pile (resp. face) ".
Lors qu'on lance une pièce, on a une chance sur 2 qu'elle tombe sur pile.
Avec 2 lancés successifs, le résultat peut être :
PP ; PF ; FF ; FP. Soit 4 possibilités, la probabilité de chaque cas est donc ¼.
Avec 3 lancés successifs, le résultat peut être :
PPP ; PPF ; PFF ; FFF ; FPP ; FPF ; FFP ; PFP. Soit 8 possibilités. Etc.
La probabilité de suite continue de 1 pile (resp. face) (ie changement) est 1/2
La probabilité de suite continue de 2 pile (resp. face) est 1/4
La probabilité de suite continue de 3 pile (resp. face) est 1/8
La probabilité de suite continue de N pile (resp. face) est 1/ 2^N

Si on veut distinguer les suites de pile et les suites de face, ces valeurs sont naturellement à diviser par 2.

Il n'est pas très difficile de faire un petit programme qui simule un jet de pièce et compte le nombre de suites de 1 (ie changement), de 2 pareils, de 3 pareils etc.

Ensuite on compare les résultats observés (fréquence des évènements) aux probabilités calculées et on constate que sur un nombre assez grand, cette loi est vérifiée.
Cela autorise à faire l'affirmation suivante : "il est impossible qu'avec une pièce équilibrée, les 100 premiers jets tombent sur la même face".
La probabilité que cela arrive est égale à 1/ 2^100. (ce qui correspond, à raison de 100 jeux à l'heure, à de très nombreux milliards d'années).

Par contre, il est certain que lors d'un jeu pendant très longtemps, cette situation arrivera, puisque sa probabilité n'est pas nulle. Mais elle reste "impossible" dans le monde réel.

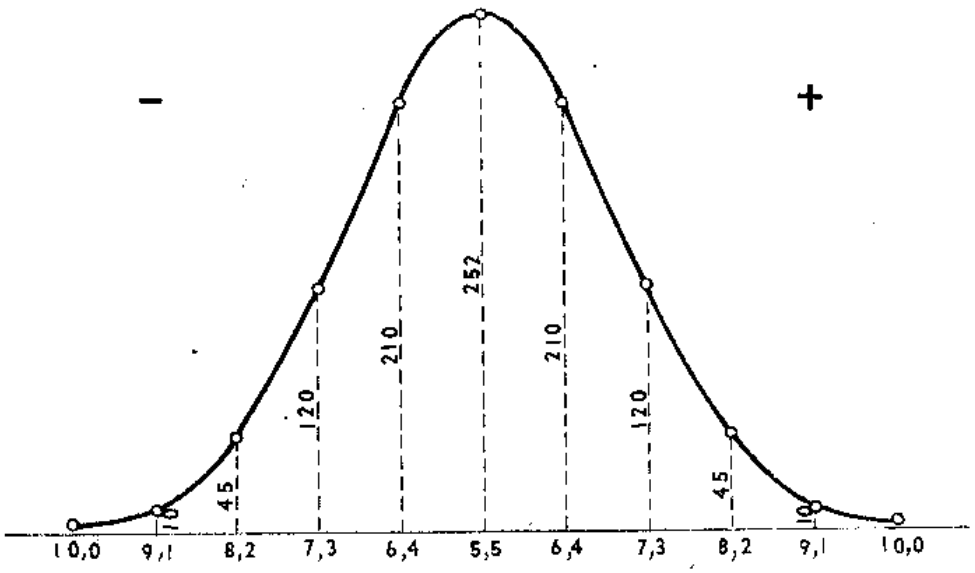
Second théorème de Bernoulli - loi Normale

Les détails dépassent le cadre de ce papier, je donnerai seulement les conclusions.

Lorsque le nombre des épreuves (jeu de pile ou face) devient infini :
1°) la combinaison la plus probable est celle qui comporte autant de coups pile que de coups face;
2°) les points représentatifs des coefficients tendent vers les points de la courbe

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Supposons en effet n très grand et p variable entre 0 et n.



On reconnaît la courbe que l'on avait déjà tracée en joignant les sommets des rectangles ou le remplissage des godets avec les billes de la planche de Galton.
Il s'agit de la courbe Gauss, représentative de la loi normale.

Le premier théorème de Bernoulli nous indique que la fréquence tend vers la probabilité et que par conséquent la moyenne arithmétique tend vers la valeur la plus probable de l'inconnue. Ceci confirme le postulat de la moyenne.

Ecart-type.

La définition de l'écart_type est la moyenne quadratique des écarts à la moyenne.

Il est bien évident que, si on peut toujours calculer une moyenne quadratique de différentes valeurs, le qualificatif d'"écart-type" ne vaut que si la répartition des écarts est conforme à la loi normale.
Dans les calculs on utilise quelque fois la variance, c'est le carré de l'écart-type. C'est le cas pour le calcul de composition des lois de probabilité.
"Quelles que soient les lois de probabilités auxquelles sont astreintes les variables éventuelles, l'erreur moyenne quadratique d'une somme est égale à la racine carrée de la somme des carrés des erreurs moyennes quadratiques composantes".
Cette notion est bien connue de ceux qui ont à évaluer les erreurs ou incertitudes des méthodes et appareils de mesures : "les erreurs accidentelles se composent quadratiquement".

Calcul de l'écart type. On dispose qu'une série de mesures ou d'observations d'une même chose. Deux situations sont possibles :
Soit la valeur vraie de la mesure de la chose est connue, alors l'écart type est la racine carrée de la somme des carrés des différences de chaque mesure à la valeur vraie, divisée par le nombre de mesures.
Soit la valeur vraie est inconnue, et c'est le cas le plus fréquent, alors la moyenne M est la moyenne arithmétique de toutes les mesures et l'écart-type est la racine carrée de la somme des carrés des différences de chaque mesure à la moyenne M, divisée par le nombre de mesures moins un.

Cela se démontre, mais il est facile de le comprendre. Supposons que l'on fasse 2 mesures d'une même chose et que la valeur vraie soit connue.
v1 et v2 les 2 mesures et A la valeur vraie.
e1 = v1 – A ; e2 = v2 – A
$\sigma = \mathrm{sqrt}((e1^2 + e2^2)/2)$.

Supposons maintenant que la valeur vraie n'est pas connue. On va calculer la moyenne M.
M = (v1 + v2)/2.
e1 = v1 – M ; e2 = v2 - M
$\sigma = \mathrm{sqrt}((e1^2 + e2^2)/(2-1))$.
Après développement et simplification, on obtient
$\sigma = \mathrm{sqrt}(1/2\ (v1 - v2)^2)$ ce qui correspond bien à la loi de composition des erreurs accidentelles.

Supposons maintenant que l'on n'ait qu'une seule mesure. A l'évidence l'écart type est inconnu, c'est à dire indéterminé et certainement pas 0, que l'on obtient bien en calculant l'expression qui vaut 0/0.

On utilise généralement la valeur de l'écart-type comme unité pour évaluer et comparer la dispersion, c'est à dire la qualité des mesures et observations.

On aurait pu aussi utiliser l'écart moyen arithmétique qui est égal à environ 4/5 de l'écart moyen quadratique. La valeur exacte est rac(2/π). La démonstration dépasse le cadre de ce papier. La comparaison des deux valeurs (ema/emq ~ 4/5) peut être utilisée pour vérifier que l'expérience respecte bien la loi de Gauss.

Répartition des écarts à la moyenne.

Je ne l'ai pas précisé explicitement, mais il apparaît que la fonction de répartition des écarts à la moyenne, la loi normale, représentée par la courbe de Gauss est unique, à une translation et une affinité près. Graphiquement, on appelle cela un changement d'échelle. On parle donc de la loi normale centrée réduite.
On a établi des tables de répartition. Celles-ci fournissent des valeurs avec une précision de l'ordre de 4 chiffres significatifs (ce qui est généralement plus que suffisant).
Pour les besoins habituels, on peut utiliser la méthode suivante.
On appelle "écart probable" l'abscisse pour laquelle la demi-courbe de Gauss est partagée en 2 aires équivalentes.
L'écart probable (ep) est égal aux 2/3 de l'écart moyen quadratique, c'est à dire aux deux tiers de l'écart-type.
La répartition des fréquences des écarts, c'est à dire la répartition des aires sous la courbe de Gauss est la suivante :
25% inférieur à 1 ep
16% compris entre 1 ep et 2 ep
7% compris entre 2 ep et 3 ep
2% compris entre 3 ep et 4 ep.
La valeur de 4 ep est généralement considérée comme une limite de validité, ou de tolérance, suivant le type d'expérience, en effet il n'y a que 0.35% des écarts qui sont au-delà de cette valeur.

Dans la littérature actuelle on lit souvent "une limite à 2 σ" qui correspond approximativement à un probabilité de 95% (47.5% de chaque côté). L'unité choisie est différente, mais le résultat est le même.

Notion de modèle.

Un modèle est une formule ou un ensemble de formules plus ou moins compliqué qui permet de calculer une valeur en connaissant une ou plusieurs variables. Un modèle peut être établi de façon théorique, puis vérifié. Ce n'est pas le contexte du présent papier.

La méthode expliquée ici est basée sur la connaissance d'observations assez nombreuses, l'établissement du modèle suit le schéma inverse : la formule est le résultat final et non l'hypothèse de départ.

Pour commencer, voici un exemple de formule très simple. Il s'agit d'évaluer l'intensité maximale de la pluie d'une durée t, de fréquence de dépassement F. Les détails concernant la pluviométrie elle-même ne sont pas traités ici, mais peuvent être trouvés dans les fichiers d'aide de l'application Assainissement de ce même site.
La formule est $i(t,F) = a(F)\, t^{b(F)}$
a et b sont des paramètres appelés "coefficients de Montana".
Pour une valeur de F donnée, on a une collection de couples (i, t).
Prenons le logarithme de chaque membre
$\log(i) = \log(a) + b \log(t)$
Ceci est une équation linéaire à 2 inconnues log(a) et b. La solution de cette équation est celle qui minimise les écarts entre la valeur calculée pour i et la valeur observée, ceci pour chaque couple observé (i, t).

Ceci nous conduit à utiliser la méthode des moindres carrés.

Méthode des moindres carrés.

"On peut la justifier à divers points de vue en généralisant la notion de moyenne pour les fonctions de plusieurs variables.
Soit un ensemble de mesures ou d'observations $x_1\ x_2\ \ldots\ x_n$.
La moyenne arithmétique de ces n valeurs rend minimum la somme des carrés des résidus.
Nous avons en effet :

$$x - x_1 = v_1$$
$$x - x_2 = v_2$$
$$x - x_n = v_n$$

La somme des carrés des résidus est égale à :

$$(x - x_1)^2 + (x - x_2)^2 + \ldots + (x - x_n)^2 = \Sigma v^2$$

Elle sera minimum si la dérivée est nulle, c'est à dire si :

$$2\,[(x - x_1) + (x - x_2) + \ldots + (x - x_n)] = 0$$

c'est la valeur de la moyenne arithmétique. On généralise le raisonnement :

Puisque pour la moyenne arithmétique la somme des carrés des résidus est minima, on conviendra d'appliquer le principe, quelle que soit la forme des relations d'observations. On démontre d'ailleurs que cette solution est la « plus probable » au sens du calcul des probabilités."

Application à la formule d'intensité de pluie.

L'équation à résoudre est
$Y = A + B.X$
C'est la forme générale lorsque l'on dispose de couples (X,Y) mesurés. A et B sont les paramètres à déterminer, ce sont donc les inconnues de l'équation.
Pour chaque couple mesuré, on peut écrire $Y = A + B.Xi$.
L'écart sur le résultat $Ei=((A + B.Xi) - Yi)$
Ecrivons que la somme des carrés de ces écarts est minimum,
$\Sigma Ei^2 = \Sigma [((A + B.Xi) - Yi)^2]$, pour I de 1 à n
Après développement et calcul de la dérivée par rapport à A et par rapport à B, on obtient le système suivant :
$nA + B.\Sigma Xi = \Sigma Yi$
$A.\Sigma Xi + B.\Sigma Xi^2 = \Sigma XiYi$
On en déduit les valeurs de A et B, coefficient de la fonction représentant le modèle.

Ceci est un calcul simple, puisqu'il n'y a que 2 paramètres, d'une méthode tout à fait générale. Elle est connue sous le nom de régression linéaire. A noter que le terme "linéaire" fait référence au degré du système à résoudre qui est 1, et non pas à la forme apparemment linéaire de la fonction représentant le modèle.

La formule du modèle de Caquot (voir détails dans la partie Assainissement) est un exemple intéressant à plusieurs titres :

- elle a été officialisée en 1977 et apparemment n'a pas été remise en cause
- elle concerne indirectement tout le monde, puisque chacun est concerné par la pluie,
- l'impact financier est considérable, d'une part puisqu'il concerne les ouvrages à créer à titre prévisionnel, d'autre part les coûts des dégâts causés par des événement de fréquence faible sont très importants.

<u>Les régressions.</u>

J'ai détaillé le calcul d'une régression avec un changement de variable en utilisant le logarithme.
Suivant que l'on applique ce changement de variable sur l'une des variables ou sur les deux, on peut facilement calculer des régressions linéaires suivant les 4 fonctions suivantes :
Régression suivant une droite : $y = a + bx$

Ajustement suivant une courbe exponentielle : $y = a\ e^{bx}$
Ajustement suivant une courbe logarithmique : $y = a + b\ \ln(x)$
Ajustement suivant une courbe de fonction puissance : $y = a\ x^{b}$

Tout ceci est une application courante des notions de base des probabilités. Je voulais l'évoquer pour montrer, à titre d'exemple, l'importance de tout cela.

Il y a des problèmes de régression beaucoup plus compliqués que ceux évoqués ici, mais il sont toujours basés sur le même principe : minimiser les écarts résiduels.

Prenons un exemple concret : on cherche à évaluer la durée de vie d'ampoules d'un certain type.
On sait que celle-ci suit une loi exponentielle, c'est à dire que plus la lampe est ancienne, plus elle a de chances de claquer. Une loi exponentielle a une médiane, c'est à dire la limite telle que la moitié des résultats lui sera inférieure, l'autre moitié lui sera supérieure. On sera exactement dans le contexte du jeu à pile ou face.

Conclusion.

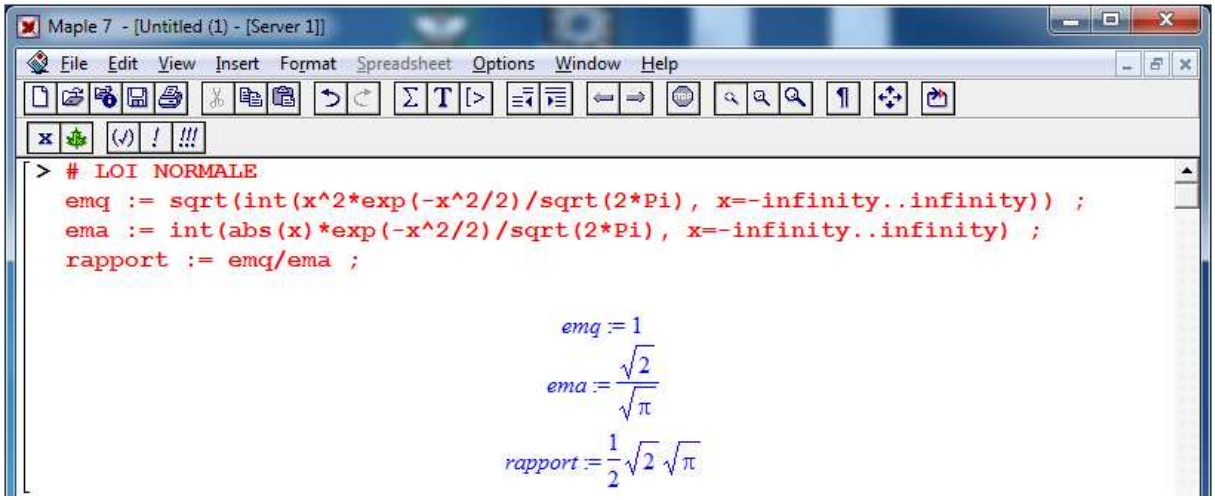
Voici deux exercices très différents mais qui représentent des cas réalistes.

Exercice 1.
Je suis face à un problème d'optimisation :
Mon client me dit que je dois livrer la pièce de remplacement de la pièce défectueuse au bout de 30jours.
Moi je ne peux livrer la pièce qu'au bout de 70jours, sachant que par an j'ai 27 demandes, quel est le stock minimum que je dois mettre en place pour pouvoir assurer mon engagement comment formaliser ce problème mathématiquement.
(Copie d'une question posée sur un forum).

Exercice 2
Un entreprise de pêche dispose de plusieurs chalutiers.
Pour des raisons évidentes de simplification cette entreprise a passé un contrat avec divers clients qui consiste à fournir dès le retour au port une ou plusieurs caisses.
On sait que les poissons les plus appréciés sont ceux qui sont ni trop petits, ni trop gros.
Le directeur de l'entreprise soupçonne certains patrons pêcheurs de réserver aux caisses marquées "Client" les meilleurs poissons. Comment peut-il le vérifier et le prouver ?.

Pierre Dolez (24 septembre 2014)

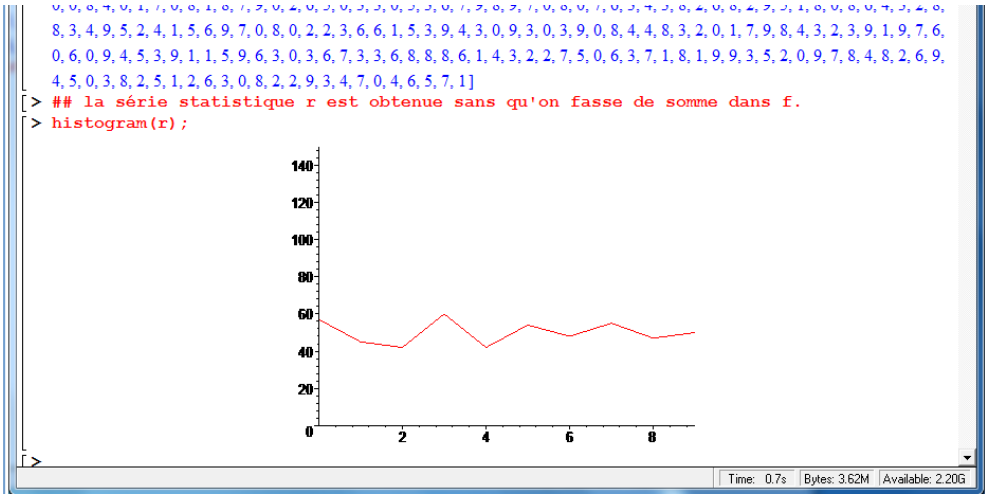
Vérification du rapport emq/ema



Vérification et illustration du TCL

A titre d'information, voici une partie de la simulation que je n'ai pas jugé utile de joindre à l'époque, mais que je rajoute (30/08/2017). Ces échanges ont été faits par courrier électronique le 18/09/2015.

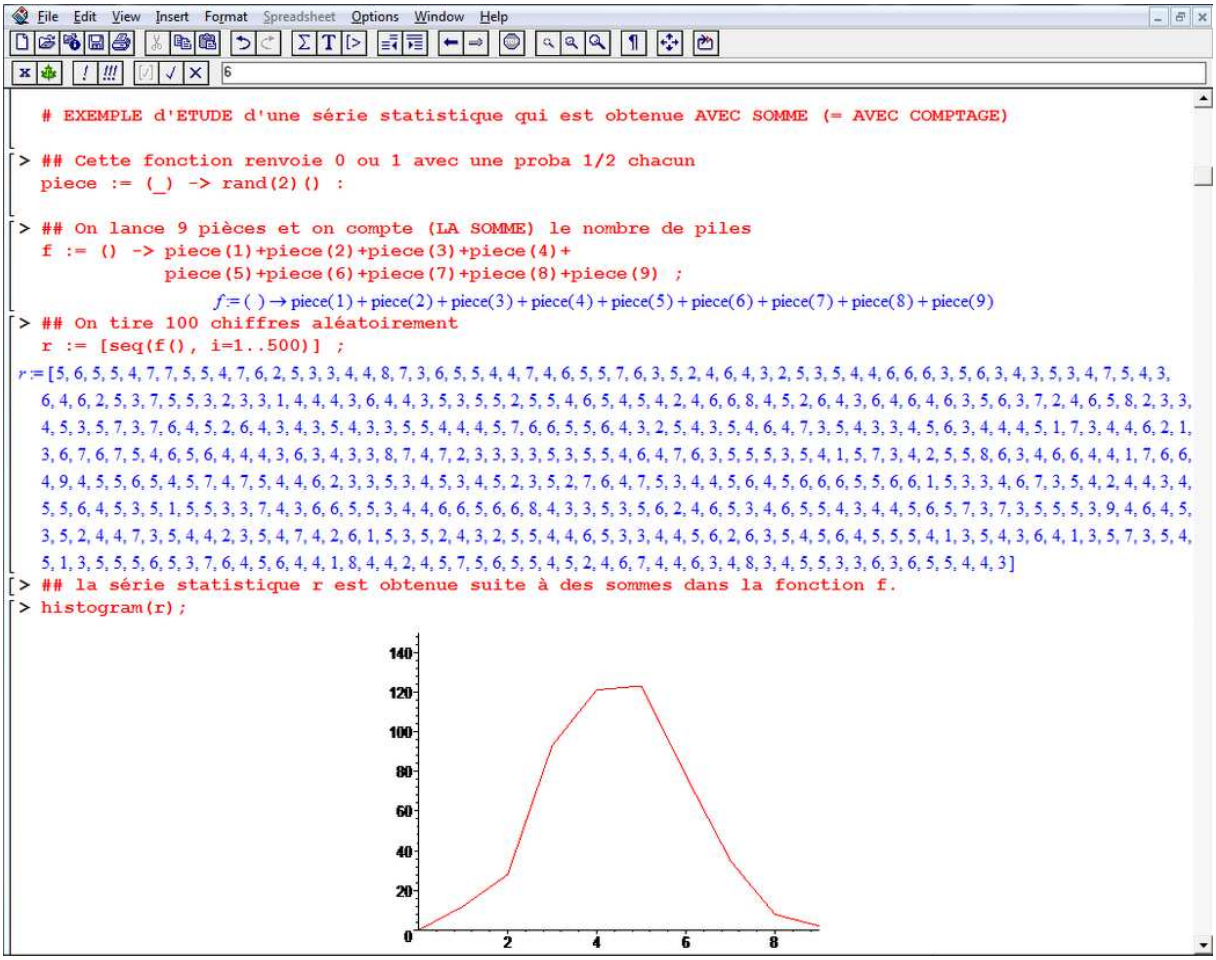
Ca va reprendre des choses que tu as déjà faites, c'est bien ainsi.
Je vais surtout insister sur les lois suivies par la fonction f, la fonction essentielle de l'histoire... tu vas voir.
La première constatation consiste tout simplement à tracer l'histogramme d'une série statistique
de 500 chiffres (compris entre 0 et 9) tirés aléatoirement avec la fonction rand() , qui est sensée suivre la loi uniforme.
Dans cette optique, ci-dessous, f est une fonction qui donne un nombre aléatoire entre 0 et 9 :
la loi suivie par f est la loi uniforme puisque, comme tu le vois, la fonction f utilise seulement rand(10).
Par ailleurs, r est un tableau qui stocke 500 tirages (donnés par f), et on trace l'histogramme donné par le tableau r :



Ci-dessus, l'histogramme est plutôt uniforme :
ce n'est pas étonnant car la loi de probabilité suivie par la fonction f est la loi uniforme entre 0 et 9.
Là, rien de génial, il n'y a pas de lien du tout avec le TCL.

Cela consiste à tracer l'histogramme d'une série statistique de 500 chiffres (compris entre 0 et 9) qui sont le nombre de fois qu'on a obtenu Pile (probabilité = 0.5) avec 9 lancés de pièces.

Dans cette optique, ci-dessous, la fonction "piece" simule un lancé de pièce, et renvoie 0 si on voit (virtuellement) face, 1 si on voit (virtuellement) pile ;
La fonction f est une fonction qui retourne un nombre compris entre 0 et 9 :
Mais cette fois, la fonction f est **une somme** de 9 fonctions élémentaires (= piece(1) ... piece(9))
Chaque piece(i) suit la loi uniforme sur les entiers 0 et 1 (proba = 0.5 chacun), mais la somme piece(1) + ... + piece(9) ne suit pas la loi uniforme : la loi suivie par cette somme (que la fonction f calcule) suit la loi binomiale B(9, 0.5), qui est assez proche d'une loi normale.
Enfin, r est un tableau qui stocke 500 tirages (donnés par f), et on trace l'histogramme donné par le tableau r :



Ci-dessus, l'histogramme est une courbe en cloche : c'est ce qu'annonce le TCL :
la loi normale est proche de la loi de probabilité suivie par une **somme** de plusieurs variables qui suivent une même loi uniforme
(ici, ce sont les piece(i) qui suivent la loi uniforme sur {0,1}, et f est la somme des piece(1) à piece(9))

Et la conclusion (rajouté le 30/08/2017)

L'explication mathématique de cette seconde constatation est donnée par Levallois :
dans son cours, on lit une introduction à la loi binomiale page 143 (avec les << coefficients binomiaux du binôme de Newton >>)
et le fait que la loi binomiale soit proche de la loi normale est énoncé
dans le point 2. du théorème de Bernoulli page 144 (<< les coefficients tendent vers les points de la courbe gaussienne >>)

Dans ton document, dans le paragraphe << Dispersion des écarts à la moyenne >>,
c'est la première constatation que tu présentes (à soustraction près des valeurs par leur moyenne, ce qui ne change rien en fait).

L'auteur de ces simulations n'a pas souhaité que je le nomme.