

4. Variables aléatoires et lois usuelles

4.1 Histoire

La notion de variable aléatoire est née en même temps que le calcul des probabilités sans toutefois être repérée comme telle. C'est dans le cadre de la théorie des erreurs de mesure, en Astronomie et Géodésie que les savants du XVIIIème siècle ont découvert les principales propriétés d'une variable aléatoire : comment les erreurs se distribuent-elles entre valeurs négatives et valeurs positives, entre petites et grandes valeurs ? Quel milieu d'erreur peut on choisir qui fasse au mieux coïncider observations erronées et vraie grandeur du phénomène mesuré ?

La loi binomiale a joué un rôle central dans les débats sur la loi des grands nombres c'est à dire la convergence des fréquences vers une limite appelée probabilité. Newton, Pascal, Jacques Bernoulli ont contribué à sa connaissance. Mais ce sont de Moivre, Laplace et Gauss qui ont vers fondé la théorie analytique des probabilités, donné au début du XIXème des solutions aux questions de la théorie des erreurs, et permis l'étude de lois continues. La plus célèbre est la loi en e^{-x^2} dite loi de Laplace-Gauss qui joue le rôle fondamental en statistique mathématique de loi limite de toute somme de variables aléatoire : c'est le théorème de la limite centrale (Laplace 1810). Cette loi a été enrôlée par Quetelet comme symptôme de l'homogénéité d'une population et condition d'émergence de l'homme moyen, et par Lexis et Galton comme symptôme de l'hétérogénéité d'une population, opposant « les tarés aux génies », dixit ce dernier qui l'a rebaptisée loi « normale ».

Cette section est développée et illustrée dans le diaporama « Histoire ».

4.2 Variable aléatoire discrète

4.2.1 Définition. Loi. Fonction de répartition

Une épreuve aléatoire donne lieu à des résultats qui sont très souvent des nombres, ou auxquels on peut faire correspondre sans équivoque des nombres, c'est à dire des éléments d'une échelle de mesure ou plus généralement d'une structure algébrique ayant certaines propriétés : modalités formant des classes d'équivalence, vérifiant une relation d'ordre, et susceptibles d'opérations comme l'addition, la multiplication...

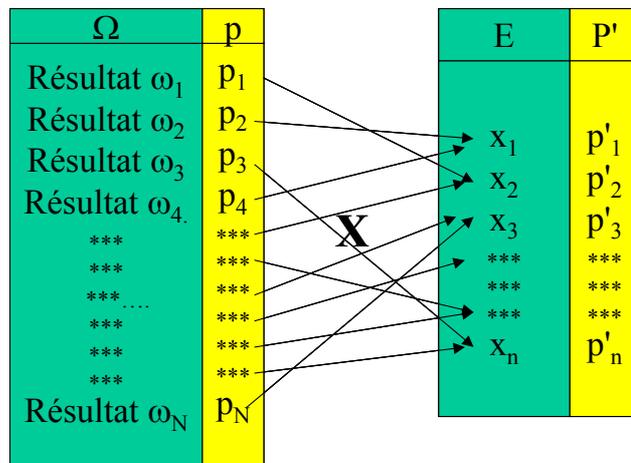
Par exemple si je lance un dé, le résultat qui m'intéresse est directement un chiffre entier entre 1 et 6 et chacune de ces valeurs a la même probabilité.

Si je lance 3 dés, et m'intéresse à la somme des points obtenus, les résultats possibles sont tous les nombres entiers entre 3 et 18, et ils ne seront pas équiprobables (cf. L3).

Si je tire une personne au hasard dans un groupe, et m'intéresse à son âge en années révolues, le résultat sera un nombre entier. Bien sûr plusieurs personnes pourront avoir le même âge, et si je tire 10 personnes, j'aurai au plus 10 valeurs de l'âge, tandis que les valeurs *possibles* seront plus nombreuses. Un résultat de l'épreuve est une personne, et chacune a par hypothèse la même probabilité d'être tirée, mais à chaque personne je ferai correspondre un âge et toutes les valeurs de l'âge n'auront pas la même probabilité.

Plus généralement une variable aléatoire X peut être considérée comme le résultat d'une application X d'un ensemble probabilisé $\{\Omega, P(\Omega), p\}$ dans un nouvel ensemble probabilisé $\{E, P(E), p\}$ où E est l'ensemble des valeurs possibles de la variable X , $P(E)$

l'ensemble des parties de E (si E est fini) ou encore des événements relatifs à X, et p' la mesure de probabilité définie sur ces événements.



En fait, grâce aux règles du calcul des probabilités, on sait que la probabilité de tout événement peut se définir à partir de la probabilité des événements élémentaires, composés d'une seule valeur de E. Il suffit donc de connaître les probabilités de ces valeurs pour connaître ce que nous appelons la loi de la variable X. Une telle loi, pour une variable aléatoire discrète, est donnée par le tableau :

Valeurs de x	x ₁	x ₂	...	x _i	...	x _n
Probabilités	p ₁	p ₂	...	p _i	...	p _n

Ce tableau traduit une fonction f qui à toute valeur x_i fait correspondre la probabilité p_i = f(x_i). Cette fonction prend le nom de densité de probabilité.

Si les valeurs x_i sont ordonnées, c'est à dire sont les éléments d'une échelle de mesure munie d'une relation d'ordre, ce qui est le cas usuel d'une variable entière, alors on peut comme en statistique définir les probabilités cumulées :

$$P(x_i) = \text{prob}(X \leq x_i) = \sum_{k=1}^{k=i} p_k$$

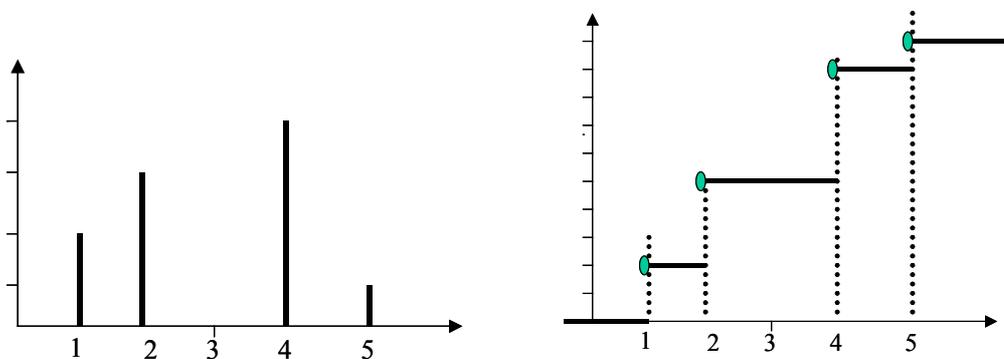
Si l'on veut répondre à la question : que vaut prob (X ≤ x) pour toute valeur de x (les valeurs observées x_i mais toute autre valeur aussi), alors on est conduit à définir la Fonction de répartition de la variable X notée F_X et qui pour la valeur quelconque x prend la valeur :

$$F_X(x) = \text{prob}(X \leq x) = P(x_i / x_i = \max\{x_k / x_k \leq x\})$$

C'est la valeur de la probabilité cumulée jusqu'à la dernière valeur possible inférieure ou égale à x. Pour une variable discrète cette fonction est discontinue (à gauche) aux points x = x_i. On peut représenter ces deux fonctions - densité et répartition – par deux graphiques, le premier sous forme de bâtons, le second sous forme d'escalier. Soit par exemple la petite loi suivante (3 n'est pas une valeur possible) :

x	1	2	4	5
p	0,2	0,3	0,4	0,1

Les fonctions densités et répartitions sont les suivantes :



4.2.2 Espérance et Variance

Comme nous avons défini la moyenne et la variance d'une variable statistique dont les valeurs étaient observées avec certaines fréquences, nous pouvons définir la moyenne théorique et la variance théorique d'une variable aléatoire dont les valeurs possibles ont certaines probabilités.

La moyenne théorique prend le nom *d'espérance mathématique*, du latin *expectatio*, que l'on peut traduire aussi bien par espérance que par anticipation. Elle représente la valeur certaine équivalente à la « loterie » que traduit la variable aléatoire. Elle en est le « juste prix » comme le disaient les inventeurs du calcul des probabilités dans le cadre des *contrats aléatoires* de type vie espérée, prime d'assurance, rente viagère, etc... Formellement c'est une simple moyenne des valeurs de la variable, pondérées par leurs probabilités :

$$E(X) = \sum_i p_i x_i$$

L'espérance est, comme la moyenne, un opérateur linéaire : si je multiplie toutes les valeurs de x par a , l'espérance est multipliée par a ; si j'ajoute b à toutes les valeurs de x , l'espérance est augmentée de b :

$$E(aX) = a.E(X)$$

$$E(X+b) = E(X) + b \quad \Rightarrow \quad E(aX+b) = a.E(X)+b$$

Mais il n'est pas vrai en général que $E[f(X)] = f[E(X)]$. Par exemple $E(X^2) \neq [E(X)]^2$. La différence entre ces deux quantités définit d'ailleurs la variance.

La variance de X se définit en effet comme moyenne (théorique) des carrés des écarts à la moyenne théorique :

$$V(X) = E[X - E(X)]^2 = \sum_i p_i (x_i - E(X))^2$$

Mais en développant le carré on aboutit à une formule plus simple (moyenne des carrés moins carré de la moyenne) :

$$V(X) = E[X^2] - [E(X)]^2 = \sum_i p_i x_i^2 - \left(\sum_i p_i x_i \right)^2$$

$\sigma_X = \sqrt{V(X)}$ est l'écart-type de la variable X .

Variance et écart-type vérifient les propriétés suivantes :

$$V(aX) = a^2.V(X) \quad \text{et} \quad \sigma_{aX} = a\sigma_X$$

(Si je multiplie les valeurs par a la variance est multipliée par a^2 et l'écart-type par a .)

$V(b) = 0$ et $\sigma_b = 0$ (une constante, cela ne varie pas)

$V(aX+b) = a^2V(X)$ et $\sigma_{aX+b} = a\sigma_X$

4.2.3 Inégalité de Bienaymé-Tchebycheff

Etablie vers 1850, cette formule donne la probabilité qu'une variable aléatoire soit dans un certain intervalle quand on ne connaît que sa moyenne et sa variance :

$$\text{Prob}\{ |X - E(X)| \leq t\sigma \} \geq 1 - 1/t^2$$

Cette probabilité est toujours sous-évaluée et si on connaît la loi de X le calcul exact aboutit à des valeurs assez différentes :

Prenons par exemple une loi uniforme (cf. 4.3.3.) entre 0 et 1 : $m = E(X) = 0,5$ et $V(X) = 1/12$, donc $\sigma = 1/2\sqrt{3}$.

Pour $t = 3$ on obtient $3\sigma = \sqrt{3}/2 = 0,866$ et l'inégalité de BT donne :

$\text{Prob}\{ |X - 0,5| \leq 0,866 \} \geq 8/9$ mais cette probabilité est égale en fait à 1.

Pour $t = 1,5$ on obtient $3\sigma/2 = \sqrt{3}/4 = 0,433$ et l'inégalité de BT donne :

$\text{Prob}\{ |X - 0,5| \leq 0,433 \} \geq 5/9 = 0,555$

Mais en fait $\text{prob}\{ |X - 0,5| \leq 0,433 \} = F(0,933) - F(0,067) = 0,866$.

L'inégalité ne donne qu'une estimation par défaut très grossière de cette probabilité.

4.2.4 Couple de V.A. discrètes. Indépendance

Supposons que l'on définisse deux variables aléatoires X et Y sur la même épreuve aléatoire. Par exemple si nous lançons deux dés, un résultat est le couple (X1,X2), et nous pouvons définir X = somme des points = X1 + X2 ; Y = différence absolue des points = |X1 - X2|. On encore nous tirons des personnes adultes au hasard dans une population et nous mesurons X = nombre d'enfants et Y = nombre de frères et soeurs.

Pour chaque résultat de l'épreuve aléatoire, le couple (X,Y) prend des valeurs (xi, yi) et de la probabilité des résultats nous pouvons déduire la probabilité de ces couples, c'est-à-dire :

$p_{ij} = \text{prob} [(X=x_i) \text{ et } (Y=y_j)]$. La loi du couple (X,Y) est donnée par un tableau à double entrée où l'on trouve p_{ij} à l'intersection de la ligne i et de la colonne j, comme on l'a vu pour deux dés discernables dans la leçon 3. De ce tableau on peut déduire très facilement la loi de toute nouvelle variable Z fonction de X et Y. Par exemple pour le jet de 2 dés nous obtenons la loi de la somme S par le biais de ce tableau des couples ordonnés (X1,X2) qui ont chacun une probabilité 1/36 :

X1 / X2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Loi de la somme $S = X_1 + X_2$:

S	2	3	4	5	6	7	8	9	10	11	12
prob	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

La loi d'un couple peut être résumée par 5 paramètres : les espérances et variances de X et Y, et la covariance du couple (X,Y) qui est définie par « moyenne des produits moins produit des moyennes » :

$$\text{Cov}(X,Y) = E[(X-E(X))(Y-E(Y))] = E(XY) - E(X)E(Y)$$

Il est facile de montrer que :

$$E(X+Y) = E(X) + E(Y)$$

En reprenant la formule de la variance, montrez que :

$$V(X+Y) = V(X) + V(Y) + 2 \text{Cov}(X,Y)$$

Deux variables X et Y sont indépendantes en probabilité si tous les événements relatifs à l'une et à l'autre sont indépendants, donc si :

$$\forall i,j, \text{prob} \{(X = x_i) \text{ et } (Y=y_j)\} = \text{prob}(X = x_i) \cdot \text{prob}(Y=y_j)$$

On peut alors montrer que dans ce cas de l'indépendance :

$$E(XY) = E(X).E(Y) , \text{Cov}(X,Y) = 0 , \text{ et } V(X+Y) = V(X) + V(Y)$$

4.2.5 Loi uniforme

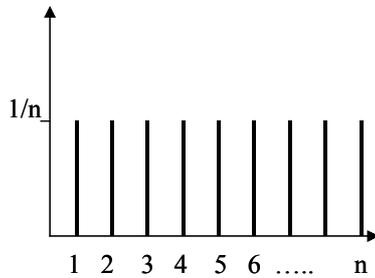
Nous passons maintenant en revue les lois discrètes les plus usuelles.

Lorsque les valeurs d'une variable ont toutes la même probabilité on parle de loi uniforme. C'est le cas par exemple de la loi des points amenés par le dé. C'est aussi le cas d'un tirage *au hasard* dans une population. En effet cette expression « au hasard » signifie précisément que nous donnons à chaque individu de la population la même chance d'être tiré, soit $1/N$ si N est la taille de cette population.

Il n'est pas si facile que cela de tirer des individus au hasard. On a besoin pour cela d'une part d'une liste complète et numérotée de ces individus, d'autre part d'un processus de génération de nombres au hasard. C'est si délicat à produire physiquement que les ingénieurs et statisticiens ont produit dans les années 1940-1950 (par exemple à la RAND Corporation) des tables de nombres au hasard que l'on peut utiliser. Aujourd'hui on se contente d'un générateur de nombres au hasard programmé sur l'ordinateur. Tout langage de programmation en inclut un. En fait il fournit des pseudo nombres aléatoires puisqu'ils sont issus d'un procédé déterministe (une suite d'opérations arithmétiques).

La forme commune de la loi uniforme entre 1 et n, notée $U(n)$ est la suivante (on pourrait la généraliser comme loi uniforme entre a et b en créant la variable $Y = X+a-1$) :

X	1	2	3	...	k	...	n
p	1/n	1/n	1/n	...	1/n	...	1/n



$$\text{Prob}(X=k) = 1/n$$

$$E(X) = (n+1)/2$$

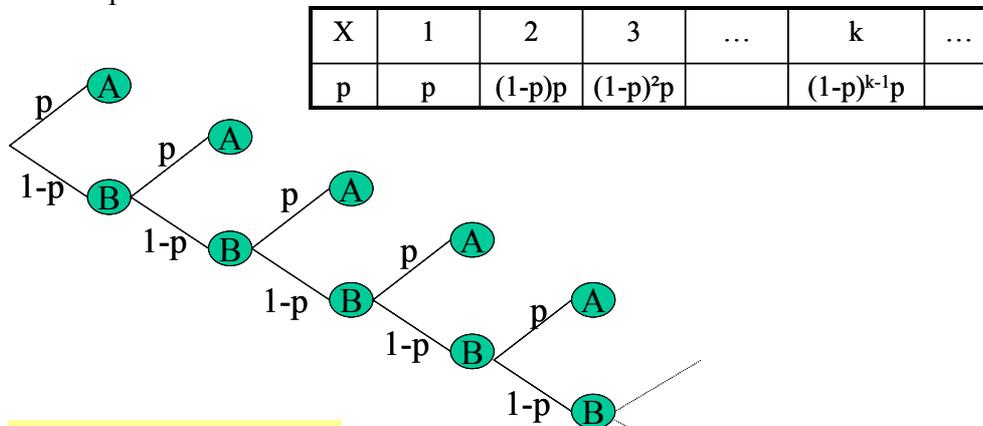
$$V(X) = (n^2-1)/12$$

Exemple : tirage au hasard dans une urne avec n boules numérotées

4.2.6 Loi géométrique

Dans une population, on a une proportion p d'individus A et $(1-p)$ d'individus B. On tire au hasard avec remise jusqu'à ce que l'on obtienne un individu A. Ou bien encore on répète une épreuve quelconque dont les deux issues sont A (probabilité p) et B (probabilité $(1-p)$) jusqu'à ce que l'on obtienne A. X = nombre de tirages nécessaires suit une loi géométrique de paramètre p notée $G(p)$.

L'arbre des possibilités est le suivant :



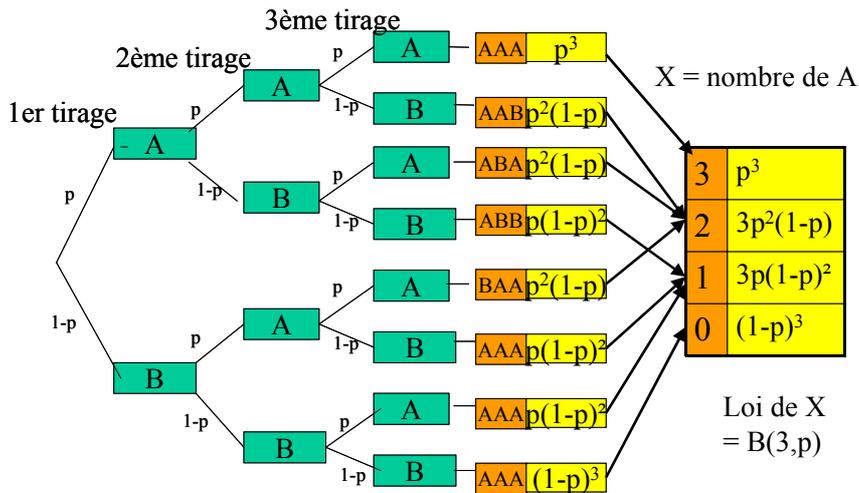
$\text{Prob}(X=k) = (1-p)^{k-1}p$
 $E(X) = 1/p$
 $V(X) = (1-p)/p^2$

Remarque : les valeurs possibles de la variable sont les valeurs entières de 1 à l'infini. Mais l'espérance et la variance sont finies

4.2.7 Modèle binomial

Le point de départ est le même que précédemment : dans une population, on a une proportion p d'individus A et $(1-p)$ d'individus B. On tire encore au hasard, mais cette fois-ci, on répète n fois ce tirage avec remise. Ou bien encore on répète n fois indépendamment une épreuve quelconque dont les deux issues sont A (probabilité p) et B (probabilité $(1-p)$). X n'est plus le nombre de tirages (fixe = n) mais le nombre d'individus (ou d'issues) A obtenus dans l'échantillon de taille n . X suit une loi Binomiale de paramètre n et p notée $B(n,p)$.

Dans le cas $n=3$, l'arbre des possibilités est le suivant :



Plus généralement c'est un arbre à 2^n branches qui génère autant d'échantillons ordonnés possibles. Chacun d'eux comprend un nombre X de résultats « A ». Ce nombre prend toutes les valeurs entières entre 0 et n . La probabilité associée à la valeur k est égale à la probabilité d'un résultat comprenant k fois A soit $p^k(1-p)^{n-k}$ que multiplie le nombre de chemins (d'échantillons) comprenant k fois A. Ce nombre est le nombre de façons de choisir k tirages parmi n soit le nombre de combinaisons C_n^k . Finalement la loi d'une variable binomiale $B(n,p)$ se résume par ce tableau :

X	0	1	...	k	...	n
prob	$(1-p)^n$	$p(1-p)^{n-1}$		$C_n^k p^k (1-p)^{n-k}$		p^n

X_i	1	0
Prob	p	1-p

On pourrait calculer l'espérance à partir de ce tableau. Mais le calcul est beaucoup plus simple en considérant que X est la somme de n variables de Bernoulli X_i qui valent 1 au i ème tirage si on a obtenu A, et 0 sinon (le nombre total de A dans l'échantillon est bien la somme de ceux que j'ai obtenus à chaque tirage). Or la loi de cette variable de Bernoulli X_i (ci-dessus) conduit à :

$$E(X_i) = 1.p + 0.(1-p) = p$$

$$E(X_i^2) = 1.p + 0.(1-p) = p$$

$$V(X_i) = E(X_i^2) - [E(X_i)]^2 = p - p^2 = p(1-p)$$

Donc puisque la variable binomiale est une somme de n variables bernoulliennes indépendantes de ce type, nous avons :

$$E(X) = np$$

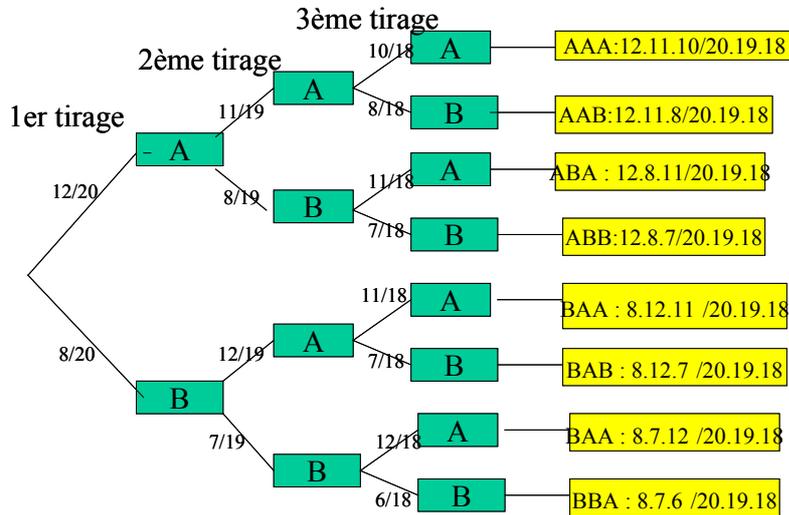
$$V(X) = np(1-p)$$

La somme de deux variables binomiales indépendantes $B(n_1,p)$ et $B(n_2,p)$ est encore une variable binomiale $B(n_1+n_2,p)$.

Le calcul des probabilités binomiales étant un peu fastidieux, nous verrons que l'on peut très souvent utiliser des lois approchées tabulées pour y remédier.

4.2.8 Modèle hypergéométrique

C'est le même modèle que celui de la binomiale à ceci près que l'on opère un tirage *sans remise* ce qui a pour conséquences que la probabilité va changer à chaque tirage en fonction des tirages précédents, et que ceux-ci ne sont plus indépendants. Commençons par un exemple avec un tirage de 3 individus dans une population comprenant 12 « A » et 8 « B » :



Plus généralement nous devons considérer le tirage de n individus dans une population de N individus dont la proportion p a la caractéristique « A ». Le nombre de « A » dans l'échantillon suit alors une loi hypergéométrique à 3 paramètres notée $H(N,n,p)$.

La loi de cette variable est :

X	0	1	...	k	...	n
Prob	$\frac{A_{N(1-p)}^n}{A_N^n}$			$\frac{C_{Np}^k C_{N(1-p)}^{n-k}}{C_N^n}$		$\frac{A_{Np}^n}{A_N^n}$

On peut montrer que :

$$E(X) = np$$

$$V(X) = \frac{N-n}{N-1} np(1-p)$$

On voit que ce sont presque les mêmes caractéristiques que pour la loi binomiale au facteur $(N-n)/(N-1)$ près (dit facteur d'exhaustivité). Notons que dès que n est faible devant N (disons $< 1/100$) ce facteur est proche de 1 (à 1% près) et les probabilités données par la formule hypergéométriques sont très proches de celles qui sont données par la formule binomiale. En pratique nous considérons donc dans ce cas que $B(n,p)$ est une loi approchée de $H(N,n,p)$ qui en offre une bonne approximation.

4.2.9 Loi de Poisson

Cette loi découverte par Simon-Denis Poisson, et redécouverte à la fin du XIXème par Bortkiewicz à propos d'une statistique des décès annuel dans l'armée prussienne dus à une ruade de cheval (!) est typiquement la loi d'événements rares, c'est à dire de probabilité faible : erreur de frappe, impureté dans un produit, panne, achat pour une personne touchée par une publicité...

Partons du modèle binomial et supposons que n est assez grand (disons supérieur à 30) et que p est faible (disons inférieur à $\frac{1}{4}$), alors on peut montrer que la probabilité binomiale tend vers celle d'une loi de Poisson de paramètre $\lambda = np$, notée $P(\lambda)$. Comme pour la loi géométrique, les valeurs possibles d'une variable de Poisson sont les nombres entiers de 0 à l'infini, mais bien sûr ces probabilités décroissent assez vite pour devenir négligeables. Les caractéristiques de la loi de Poisson $P(\lambda)$ sont les suivantes :

X	0	1	...	k	...	∞
Prob	$e^{-\lambda}$	$\lambda e^{-\lambda}$		$p_k = \frac{\lambda^k e^{-\lambda}}{k!}$		0

Cette loi est tabulée sous la forme d'un tableau de ce type pour différentes valeurs de λ et sous la forme plus pratique de probabilités cumulées pour les mêmes valeurs de λ . Un calcul classique de somme en série d'une infinité de termes montre que la somme des p_k vaut 1 et que :

$$E(X) = \lambda$$

$$V(X) = \lambda$$

La somme de deux variables de Poisson indépendantes de paramètres λ_1 et λ_2 est encore une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.

4.3 Variable aléatoire continue

4.3.1 Définition. Densité et fonction de répartition. Moments.

La considération de variables aléatoires à valeurs dans l'ensemble continu des nombres réels pose quelques difficultés théoriques pour définir l'espace probabilisable et pour mener à bien certains calculs qui nécessitent des connaissances de calcul différentiel. Faisons simple en sacrifiant la rigueur à la pertinence du raisonnement.

Reprenons les exemples de variable aléatoire du 4.2.1. Si je tire une personne au hasard dans une population, et m'intéresse à l'âge exact de la personne, ou à sa taille, ou à son revenu annuel, la variable n'aura plus un nombre fini de valeurs possibles discrètes (entières dans le cas de l'âge en années révolues) mais un nombre infini (en théorie) ou innombrable (en pratique) de valeurs appartenant à une échelle continue, parce que la grandeur mesurée est elle-même continue, et que seule la précision limitée de l'instrument de mesure transforme le nombre réel de la mesure en nombre décimal (par exemple 1,82m pour la taille). Le modèle théorique de l'ensemble E des valeurs possibles de la variable X est maintenant l'ensemble \mathbb{R} des réels (qui comprend, rappelons-le, les entiers, les rationnels, et les irrationnels), et qui est tel que entre 2 nombres quelconques il en existe toujours une infinité.

Cette propriété du continu interdit de procéder à une définition de la variable aléatoire en termes de valeurs et probabilité de ces valeurs car les valeurs sont en nombre infini et la probabilité d'une valeur possible (infiniment précise) est nulle.

Ce qui n'est pas nul est la probabilité que la variable appartienne à un certain intervalle de valeurs de type $[a, b]$ ou $[a, \infty)$ ou $]-\infty, b]$. On privilégie en fait le dernier type qui permet le calcul de toute autre probabilité.

(Notons qu'il n'est plus nécessaire de faire attention à distinguer $<$ et \leq puisque la probabilité d'une valeur est nulle.)

On peut donc entièrement définir la loi d'une V.A. continue par la donnée de sa Fonction de répartition (FR) :

$$F_X(x) = \text{prob}(X < x) = \text{prob} \{X \in]-\infty, x]\}$$

C'est une fonction nécessairement non décroissante qui varie de 0 à 1.

Elle permet de calculer la probabilité de tout intervalle $[a, b]$:

$$\text{Prob} [a < X < b] = \text{prob} [X < b] - \text{prob} [X < a] = F_X(b) - F_X(a)$$

Si cet intervalle est infiniment petit de la forme $[x, x+dx]$ nous obtenons :

$$\begin{aligned} \text{Prob} [x < X < x+dx] &= \text{prob} [X < x+dx] - \text{prob} [X < x] = F_X(x+dx) - F_X(x) \\ &= [(F_X(x+dx) - F_X(x))/dx] \cdot dx \end{aligned}$$

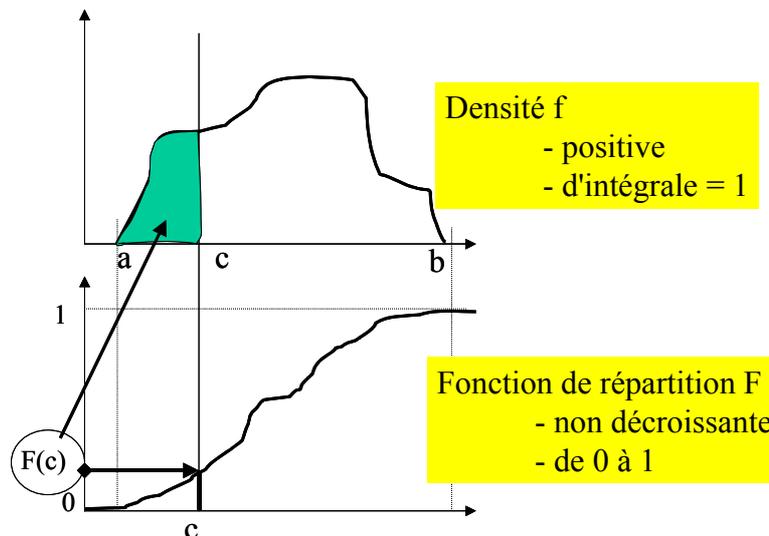
Or la quantité entre crochets, qui représente la pente de la courbe F au point x , a pour limite quand dx tend vers zéro ce qu'on appelle la dérivée de F_X en x , soit $f_X(x)$. Cette fonction f s'appelle la densité de probabilité. Comme dérivée d'une fonction croissante, elle est toujours positive. Une seconde façon de définir une V.A. continue est donc de se donner sa densité. La relation précédente s'écrit à la limite :

$$\text{Prob} [x < X < x+dx] = f_X(x) \cdot dx = \text{surface du rectangle de largeur } dx \text{ et de hauteur } f_X(x)$$

Plus généralement la probabilité que la variable X soit inférieure à b apparaîtra comme la somme (intégrale) des surfaces de tels rectangles, soit encore la surface qui est sous la courbe à gauche de b . Et la surface totale sous la densité est égale à 1 puisqu'elle représente la probabilité de l'ensemble des valeurs possibles de la variable.

$$\text{Prob}(X < b) = F_X(b) = \int_{-\infty}^b f_X(x) dx$$

$$\text{Prob}(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx$$



L'espérance mathématique de X est encore définie comme moyenne des valeurs pondérée par les probabilités (la densité) mais la somme est maintenant une somme intégrale.

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$

$$V(X) = E(X^2) - [E(X)]^2$$

Pas de souci cependant : vous n'aurez pas de calcul intégral et différentiel à effectuer. La seule chose est de bien comprendre le rapport entre densité et fonction de répartition. Une ordonnée de la seconde correspond à une surface de la première.

4.3.2 Couple de V.A. continues

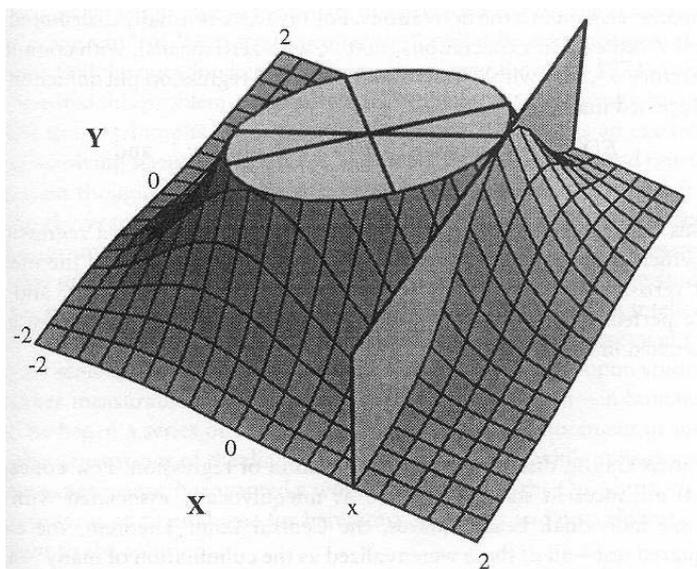
Comme pour les variables discrètes, la loi d'un couple de VA continues X, Y pourrait être donnée par un tableau croisé des valeurs possible avec dans chaque case la probabilité de ce couple. Mais le tableau aurait un nombre infini de lignes et de colonnes, et les probabilités seraient remplacées par une fonction densité qui est une fonction de deux variables $f_{X,Y}(x,y)$. Cette fonction représente maintenant non plus une courbe mais une surface dans l'espace à 3 dimensions, dont l'altitude au point de coordonnées (x,y) est $z = f(x,y)$, et tel que le volume enfermé sous cette surface soit égal à l'unité. $f(x,y)dxdy$ représentera la probabilité que X et Y soient à l'intérieur du rectangle $[x \ x+dx].[y \ y+dy]$. On peut définir des courbes de niveau de densité constante sur cette surface comme sur notre terre en faisant $z = \text{constante}$.

On peut aussi définir une partie du volume sous cette courbe qui est $F_{X,Y}(x,y) = \text{prob}\{X < x \text{ et } Y < y\}$ que l'on définira comme fonction de répartition du couple X, Y .

Si l'on projette ce volume sous la surface $f(x,y)$ sur le plan $y=0$ on obtiendra la densité marginale de X seule.

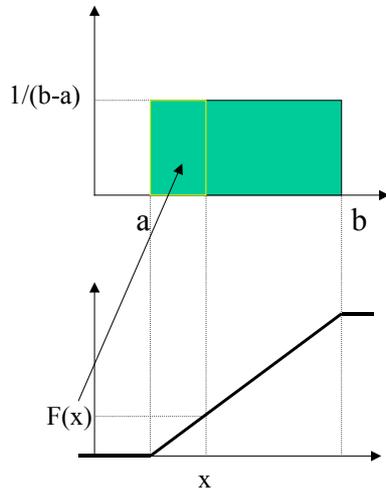
Si l'on coupe ce volume par un plan $y=k$ on obtiendra la densité conditionnelle de x pour $y=k$.

Nous n'écrirons pas les différentes formules correspondantes qui font appel à des connaissances sur les intégrales doubles. Voici une image 3D due à S. Stigler (1986) d'une loi d'un couple de variables « normale » (cf. infra).



4.3.3 Loi uniforme

La loi continue uniforme sur un intervalle $[a, b]$ est une loi de densité constante : tout intervalle élémentaire de largeur dx a la même probabilité.



$$\text{Densité : } f(x) = 1/(b-a) \text{ si } a < x < b \\ = 0 \text{ sinon}$$

$$\text{Fonction de répartition} \\ F(x) = (x-a)/(b-a) \text{ si } a < x < b$$

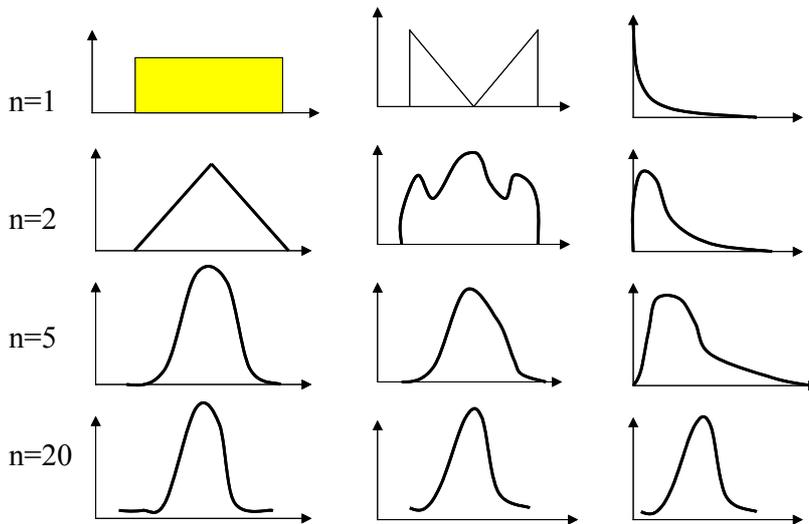
$$E(X) = (b-a)/2 \\ V(X) = (b-a)^2/12$$

4.3.4 Loi de Laplace-Gauss

Cette loi apparaît pour la première fois chez de Moivre puis chez Laplace (1774) puis chez Gauss (1800) comme loi des erreurs. Ce dernier montre en particulier que cette loi est associée au choix de la moyenne comme milieu si l'on prend le critère du maximum de vraisemblance.

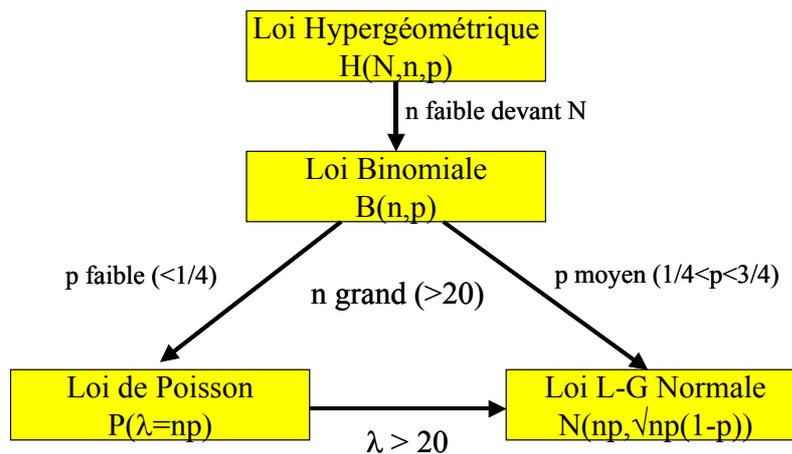
Mais c'est Laplace qui lui donne ses lettres de noblesses en démontrant en 1810 le théorème le plus fameux de la statistique mathématique : le *Théorème Central limite* (TCL) dit parfois théorème de la limite centrale. Ce théorème dit que si l'on fait la somme de n variables X_i indépendantes et de même loi, alors pour n assez grand, cette somme suit une loi de Laplace quelle que soit la loi de départ des X_i . La loi de Laplace-Gauss est donc la loi limite de toute somme (ou moyenne) de variables, du moins celles qui ont une variance finie ce que l'on a découvert par le contre-exemple de Cauchy.

Prenons par exemple des variables X_i indépendantes qui suivent la loi uniforme précédemment définie sur $[a, b]$. On peut montrer que la densité de $(X_1 + X_2)/2$ est triangulaire, puis que celle de $(X_1+X_2+X_3+X_4+X_5)/5$ est déjà plus arrondie, et que la densité de la moyenne de 20 variables X_i est une belle courbe en cloche très proche de celle d'une loi normale. Et ceci serait vrai en partant d'une autre loi, par exemple dissymétrique :



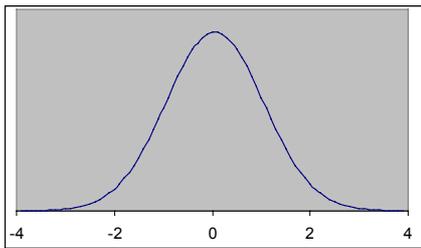
Le Théorème central limite peut s'appliquer aussi à une somme de variables de Bernoulli, donc à la loi binomiale dont la loi de Laplace Gauss est une bonne approximation si n est assez grand ($n > 20$) et p « moyen ».

Finalement on a le schéma suivant d'approximation des lois binomiale et hypergéométrique :



La loi de Laplace-Gauss, rebaptisée « loi normale » à la fin du siècle a été utilisée comme modèle de distribution « normale » des caractéristiques humaines. Non sans critiques.

Cette loi dépend de deux paramètres, m et σ , respectivement moyenne théorique (ou espérance) et écart-type. Elle est notée $N(m, \sigma)$. Il est pratique de se ramener toujours à la loi normale centrée (c'est à dire de moyenne 0) et réduite (c'est à dire d'écart-type 1) que l'on note $N(0,1)$ et qui est la seule à être tabulée. Voici ses propriétés :



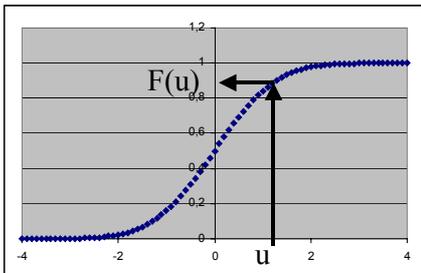
Densité :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Fonction de répartition

$$F(u) = \int_{-\infty}^u f(x) dx = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Elle est tabulée



La table (cf Infra) donne $F(u) = \text{prob}(X_{N(0,1)} < u)$ pour $u > 0$. Il faut utiliser la symétrie de la densité pour calculer certaines probabilités. Exemple ; pour $u = 1,23$ la table donne $F(u) = 0,8997 \approx 0,90$. On en déduit :

- $\text{prob}(X_{N(0,1)} > 1,23) = 0,10$
- $\text{prob}(X_{N(0,1)} < -1,23) = 0,10$
- $\text{prob}(-1,23 < X_{N(0,1)} < 1,23) = 0,80$

La lecture inverse de la table s'apparente à une table de la fonction inverse F^{-1} et permet connaissant $p = F(u)$ de trouver $u = F^{-1}(p)$. Par exemple, $p = 0,975$ correspond à $u = 1,96$.

Enfin il est possible de répondre à toute question relative à une variable normale $N(m, \sigma)$ en centrant (retirer la moyenne) et réduisant (diviser par l'écart-type) la variable pour se ramener à $N(0,1)$. Par exemple calculer $\text{prob}(Y > 5)$ sachant que $Y \rightarrow N(3,2)$. Y est une variable plus dispersée et décentrée puisque sa moyenne est 3. Mais $(Y-3)/2$ est une variable normale centrée réduite et nous nous y ramenons en faisant les mêmes opérations des deux côtés du signe ">" :

$$\text{Prob}(Y > 5) = \text{prob}\left\{\frac{(Y-3)}{2} > \frac{(5-3)}{2}\right\} = 1 - \text{prob}(X_{N(0,1)} < 1) = 1 - F_{N(0,1)}(1) = 1 - 0,8413 = 0,1587$$

4.3.5 Loi du Chi-2

C'est la loi d'une quantité qui se présente comme une somme de carrés de variables normales centrées et réduites :

Si $X_1, X_2, \dots, X_i, \dots, X_n$ sont indépendantes et suivent une loi $N(0,1)$, alors

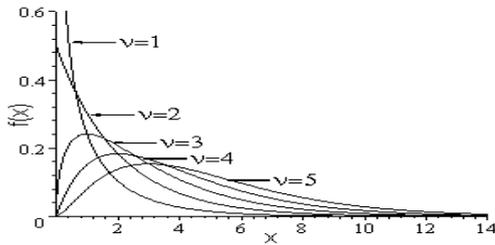
$$S = X_1^2 + X_2^2 + \dots + X_i^2 + \dots + X_n^2 \text{ suit la loi di Chi-2 : } \chi^2(n)$$

La loi du χ^2 donnée par F. R. Helmert en 1875 fut redécouverte par Karl Pearson en 1900. Elle nous sera utile un peu plus tard pour la mesure du désaccord entre l'observation et l'hypothèse (test d'adéquation du χ^2 , test d'indépendance du χ^2). Elle dépend du seul paramètre v appelé degré de liberté. Voici ses principales propriétés :

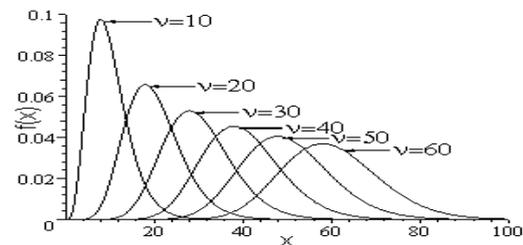
La loi du chi-deux à ν degrés de liberté, $\chi^2(\nu)$, est la loi d'une variable aléatoire X de densité :

$$f(x) = K_\nu e^{-(x/2)} x^{(\nu/2)-1}, \quad x > 0 \quad K_\nu \text{ est la constante } 1 / 2^{\nu/2} \Gamma(\nu/2)$$

Loi du chi-deux - Fonction de densité



Loi du chi-deux - Fonction de densité



-
- La loi de $X/2$ est une loi Gamma : $\gamma(\nu/2)$.
- Si $\nu = 2$, la loi $\chi^2(\nu)$ est une loi exponentielle de densité $e^{-x/2}/2$.
- $Mode(X) = \nu - 2$ si $\nu > 2$, $E(X) = \nu$ et $V(X) = 2\nu$.
- Si X et X' sont deux variables indépendantes suivant une loi du χ^2 à ν et ν' degrés de liberté, alors $X + X'$ suit une loi du χ^2 à $\nu + \nu'$ degrés de liberté.
- Si $n \geq 30$ alors $\sqrt{2X} - \sqrt{2n-1}$ suit approximativement la loi normale $N(0,1)$.
- Si $n \geq 50$, alors X suit approximativement la loi normale $N(n, \sqrt{2n})$.

4.3.6 Loi de Student

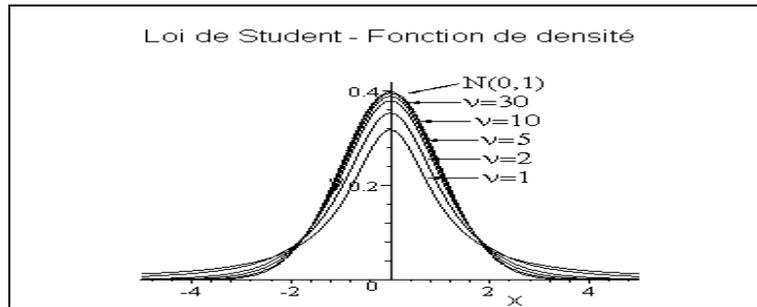
Cette loi a été mise en avant par William S. Gosset qui, travaillant dans la célèbre brasserie Guinness, a publié en 1908 sous le pseudonyme de « Student ». Elle nous sera utile pour les tests de moyenne sur petits échantillons. Elle dépend d'un seul paramètre, son degré de liberté, ν . Elle survient comme rapport d'une variable normale et de la racine d'une variable du Chi-2 :

Si Y suit $N(0,1)$ et Z suit $\chi^2(\nu)$ et qu'elles sont indépendantes,

alors $T = \frac{Y}{\sqrt{Z/\nu}}$ suit Student(ν) de densité :

$$f_T(t) = K'_\nu \left(1 + \frac{t^2}{\nu} \right)^{-\frac{\nu+1}{2}}, \quad t \in \mathbb{R} \text{ avec } K'_\nu = 1/\sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)$$

Ces courbes de densité sont en forme de cloche centrée en 0, légèrement plus évasées que pour la loi normale $N(0,1)$.



- $Mode(T) = 0$, $E(T) = 0$ si $\nu > 1$ et $V(T) = \nu/(\nu - 2)$ si $\nu > 2$.
- Si $\nu > 30$, alors T suit approximativement la loi normale $N(0,1)$.

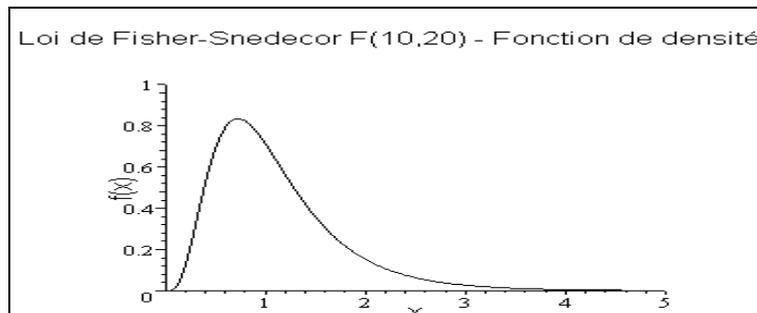
4.3.7 Loi de Fisher-Snedecor

La *loi de Fisher-Snedecor à ν_1 et ν_2 degrés de liberté*, $F(\nu_1, \nu_2)$, découverte par le statisticien anglais Ronald Fisher dans le cadre de recherches sur l'analyse de variance, est la loi d'un rapport de deux variables du Chi-2, ou plus exactement de :

$$F = \frac{\sqrt{Z_1/\nu_1}}{\sqrt{Z_2/\nu_2}} \text{ où } Z_1 ; \chi^2(\nu_1) \text{ et } Z_2 ; \chi^2(\nu_2), \text{ par exemple le rapport des variances}$$

empiriques corrigées de deux échantillons gaussiens, indépendants et de même variance. Elle a pour densité :

$$f_F(x) = K'' \frac{x^{\nu_1/2-1}}{(\nu_2 + \nu_1 x)^{(\nu_1+\nu_2)/2}}, \quad x > 0 \text{ avec } K'' = \nu_1^{\nu_1/2} \nu_2^{\nu_2/2} / B(\nu_1/2, \nu_2/2)$$



- Si $\nu_1 = 1$, alors $F_{1,\nu} = T_\nu^2$, carré d'une variable de Student.
- $1/F_{\nu_1,\nu_2} = F_{\nu_2,\nu_1}$
- $Mode(F) = \frac{(\nu_1 - 2)\nu_2}{\nu_1(\nu_2 + 2)}$ si $\nu_2 > 2$, $E(F) = \frac{\nu_2}{\nu_2 - 2}$ si $\nu_2 > 2$ et $V(F) = \frac{2\nu_2^2(\nu_2 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)(\nu_2 - 2)^2}$ si $\nu_2 > 4$.

(Pour les lois du Chi-2, de Student et de Fisher, on ne retiendra pas autre chose que leur définition et la forme de leur densité.)

Table de la Loi Normale N(0,1)

u	F(u)
0	0,5
0,1	0,5398279
0,2	0,57925969
0,3	0,61791136
0,4	0,6554217
0,5	0,69146247
0,6	0,72574694
0,7	0,75803642
0,8	0,78814467
0,9	0,81593991
1	0,84134474
1,1	0,8643339
1,2	0,88493027
1,3	0,90319945
1,4	0,91924329
1,5	0,93319277
1,6	0,94520071
1,7	0,95543457
1,8	0,96406973
1,9	0,97128351
2	0,97724994
2,1	0,98213564
2,2	0,9860966
2,3	0,98927592
2,4	0,99180247
2,5	0,99379032
2,6	0,99533878
2,7	0,99653298
2,8	0,99744481
2,9	0,99813412
3	0,99865003
3,1	0,99903233
3,2	0,9993128
3,3	0,99951652
3,4	0,99966302
3,5	0,99976733
4	0,99996831

Nous ne donnons que la table de la loi Normale.

Les autres tables peuvent être générées à la demande avec les fonctions d'Excel.

(Voir *SimulI*)