Lois de Probabilité

1. Introduction

Il est toujours possible d'associer à une variable aléatoire une probabilité et définir ainsi une **loi de probabilité**. Lorsque le nombre d'épreuves augmente indéfiniment, les **fréquences observées** pour le phénomène étudié **tendent vers les probabilités** et les distributions observées vers les distributions de probabilité ou loi de probabilité.

Identifier la loi de probabilité suivie par une variable aléatoire donnée est essentiel car cela conditionne le choix des méthodes employées pour répondre à une question biologique donnée.

2 Lois discrètes

Par définition, les variables aléatoires discrètes prennent des valeurs entières discontinues sur un intervalle donné. Ce sont généralement le résultat de dénombrement.

2.1 Loi uniforme

2.1.1 Définition

Une distribution de probabilité suit une **loi uniforme** lorsque toutes les valeurs prises par la variable aléatoire sont équiprobables. Si n est le nombre de valeurs différentes prises par la variable aléatoire, $\forall i = 1, n : P(X = x_i) = \frac{1}{n}$

Exemple :

La distribution des chiffres obtenus au lancer de dé (si ce dernier est non pipé) suit une loi uniforme dont la loi de probabilité est la suivante

X	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Avec pour espérance : $E(X) = \frac{1}{6}\sum_{i=1}^{i=6} i = \frac{1}{6}\cdot 21 \Rightarrow E(X) = 3.5$

Et pour variance : $V(X) = \frac{1}{6}\sum_{i=1}^{i=6} i^2 - E(X)^2 = \frac{1}{6}\cdot 91 - (3.5)^2 \Rightarrow V(X) = 2.92$

Où les valeurs x_i correspondent au rang i de la variable X dans la série.

2.1.2 Espérance et variance

Dans le cas particulier d'une **loi discrète uniforme** où les valeurs de la variable aléatoire

X correspondent au rang $x_i = i, (i = 1, 2, ..., n)$

Espérance	Variance
$E(X)=\frac{n+1}{2}$	$V(X)=\frac{n^2-1}{12}$

2.2 Loi de Bernoulli

2.2.1 Définition

Soit un univers Ω constitué de **deux éventualités**, S pour succès et E pour échec : $\Omega=\{E,S\}$

Sur lequel on construit une variable aléatoire discrète, « *nombre de succès* » telle que au cours d'**un**e épreuve, si S est réalisé, $X = 1$, si E est réalisé, $X = 0$

On appelle **variable de Bernoulli** ou variable ***indicatrice***,

La variable aléatoire X telle que : $X:\Omega\to\mathbb{R}$
$X(\Omega)=\{0,1\}$

La **loi de probabilité** associée à **la variable de Bernoulli** X telle que :

X	0	1
$P(X = x_i)$	q	p

avec $p+q = 1$, est appelée **loi de Bernoulli notée B(1, *p*)**

2.2.2 Espérance et variance

L'**espérance** de la variable de Bernoulli est : $E(X)=p$

La **variance** de la variable de Bernoulli est $V(X)=pq$

2.3 Loi binomiale

2.3.1 Définition

Décrite pour la première fois par Isaac Newton en 1676 et démontrée pour la première fois par le mathématicien suisse Jacob Bernoulli en 1713, la **loi binomiale** est l'une des distributions de probabilité les plus fréquemment rencontrées en statistique appliquée.

Soit l'application : $S_n:\Omega^n\to\mathbb{R}^n$

Avec : $S_n=X_1+X_2+X_3+.......+X_n$, où X_i est une variable de Bernoulli

La **variable binomiale**, S_n représente le **nombre de succès** obtenus lors de la répétition de n épreuves identiques et indépendantes, chaque épreuve ne pouvant donner que deux résultats possibles.

Ainsi la loi de probabilité suivie par **la somme de *n* variables de Bernoulli** où la probabilité associée au succès est *p*, est la **loi binomiale** de paramètres *n* et *p*. $S_n : \Omega^n \to \mathbb{R}^n$

Alors : $S_n \to B(n,p)$

La probabilité que $S_n = k$, c'est à dire l'obtention de *k* succès au cours de *n* épreuves indépendantes est :
$P(S_n = k) = C_n^k p^k q^{n-k}$

Il est facile de démontrer que l'on a bien une loi de probabilité car :

$$\sum_{k=0}^{n} P(S_n = k) = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1, car : p+q = 1$$

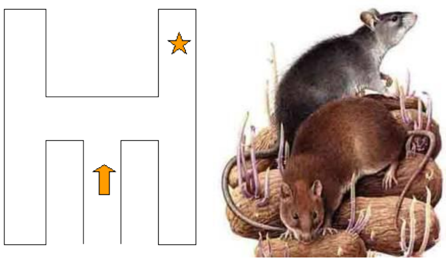
Remarque :

Le développement du <u>binôme de Newton</u> $(p+q)^n$ permet d'obtenir l'ensemble des probabilités pour une distribution binomiale avec une valeur *n* et *p* donnée.

Il existe également des tables de la loi binomiale où les probabilités sont tabulées pour des valeurs *n* et *p* données.

<u>**Exemple**</u> :

Dans une expérience sur **le comportement du rat**, *rattus norvegicus*, on fait pénétrer successivement *n* rats dans un labyrinthe en forme de H. On étudie alors la probabilité que *k* rats empruntent la branche supérieure droite du H (voir schéma ci-dessous).



A chaque épreuve, deux évènements peuvent se produire : soit le rat suit l'itinéraire voulu (succès) soit il ne l'emprunte pas (échec). Sachant qu'il y a 4 itinéraires possibles (branches), la probabilité du succès $p = 1/4$.

Hypothèse :

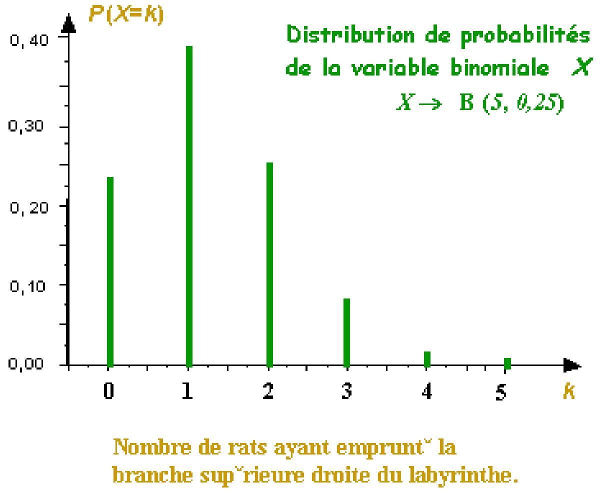
- si les rats n'ont pas été conditionnés,

- si la branche supérieure droite ne comporte aucun élément attractif ou répulsif,

- si le choix de l'itinéraire d'un rat n'affecte pas le choix du suivant (odeurs)

Alors : la variable aléatoire *X* « itinéraire emprunté pour *x* rats » suit une loi binomial : $X \to \beta\left(n, \frac{1}{4}\right)$

Dont la distribution des probabilités est la suivante si l'on étudie le comportement de **5 rats** :



k	$P(X = k)$
0	$C_5^0\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right)^0 = 0.237$
1	$C_5^1\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^1 = 0.395$
2	$C_5^2\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2 = 0.264$
3	$C_5^3\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3 = 0.088$
4	$C_5^4\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^4 = 0.015$
5	$C_5^5\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^5 = 0.001$

Remarque :

.Il est possible d'obtenir aisément les valeurs des combinaisons de la loi binomiale en utilisant le triangle de Pascal.

De plus on vérifie que la somme des probabilités est bien égale à 1.

2.3.2 Espérance et variance

L'**espérance** d'une variable binomiale S_n est égale à : $E\left(S_n\right) = np$

La **variance** d'une variable binomiale S_n est égale à : $V\left(S_n\right) = npq$

Exemple :

Dans le cadre de l'étude de comportement du rat , quel est en moyenne le nombre attendu de rats qui vont emprunter l'itinéraire prévu si l'expérience porte sur un lot de 20 rats ? Donnez également la variance et l'écart type de cette variable ?

Réponse

2.3.3 Symétrie et récurrence de la loi binomiale

La loi binomiale dépend des deux paramètres n et p. Elle est **symétrique** pour p = 0,5 et **dissymétrique** pour les autres valeurs de p. La dissymétrie est d'autant plus forte :

(1) pour n fixe, que p est différent de q

(2) pour p fixe que n est plus petit.

Afin de faciliter les calculs des probabilités, il est possible d'utiliser une formule de récurrence donnant les valeurs des probabilités successives : $P(S_n = k) = \frac{n-k+1}{k}\frac{p}{q}P(S_n = k-1)$

2.3.4 Stabilité de la loi binomiale

Théorème :

Si S_n et S_m sont deux variables indépendantes suivant des lois binomiales respectivement

$S_n \to \mathrm{B}(n,p)$ et $S_m \to \mathrm{B}(m,p)$ alors $S_n + S_m \to \mathrm{B}(n+m,p)$

2.4 Loi de Poisson

La **loi de Poisson** découverte au début du XIXe siècle par le magistrat français Siméon-Denis Poisson s'applique souvent aux phénomènes accidentels où la probabilité p est très faible ($p < 0{,}05$). Elle peut également dans certaines conditions être définie comme **limite d'une loi binomiale**.

2.4.1 Approximation d'une loi binomiale

Lorsque n devient grand, le calcul des probabilités d'une loi binomiale $P(S_n = k) = C_n^k p^k q^{n-k}$

Devient très fastidieux. On va donc, sous certaines conditions, trouver une approximation de p_k plus maniable.

Comportement asymptotique : si $n \to +\infty$ et $p \to 0$, alors $X : \mathrm{B}(n,p) \to \mathrm{P}(\lambda)$ $\quad avec : np \to \lambda$

Remarque : Cette approximation est correcte si $n \geq 50$ et $np \leq 5$

Exemple :

Soit une loi binomiale de paramètres (100 ; 0,01), les valeurs des probabilités pour k de 0 à 5 ainsi que leur approximation à 10^{-3} avec une loi de Poisson de paramètre ($\lambda = np = 1$) sont données dans le tableau ci-dessous :

k	0	1	2	3	4	5
P(X=k)	0,366	0,370	0,185	0,061	0,015	0,000
Approximation	0,368	0,368	0,184	0,061	0,015	0,003

Dans le cas de cet exemple où n =100 et np =1, l'approximation de la loi binomiale par une loi de poisson donne des valeurs de probabilités identiques à 10^{-3} près.

2.4.2 Loi de Poisson

On appelle **processus poissonnier** (ou processus de Poisson), le modèle probabiliste des situations qui voient un flux d'évènements se produire les uns à la suite des autres de façon aléatoire (dans le temps et dans l'espace), obéissant aux conditions suivantes :

- la probabilité de réalisation de l'évènement au cours d'une petite période ou sur une petite portion d'espace Δt est proportionnelle à Δt soit $p\Delta t$.
- elle est indépendante de ce qui s'est produit antérieurement ou à côté,
- la probabilité de deux apparitions sur le même Δt est négligeable.

Ainsi, des évènements qui se réalisent de façon aléatoire comme des pannes de machines, des accidents d'avions, des fautes dans un texte, …peuvent être considérés comme relevant d'un processus poissonnien.

Une variable aléatoire X à valeurs dans $\mathbb{R}$ suit une **loi de Poisson de paramètre** λ ($\lambda > 0$) si les réels p_k sont donnés par : $P(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$

On note : $X \to \mathrm{P}(\lambda)$

Remarque : Une loi de Poisson est donnée par sa loi de probabilité :

(1) $\forall k : P(X = k) > 0$

(2) $\sum_{k \geq 0} P(X = k) = \sum_{k \geq 0} \dfrac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k \geq 0} \dfrac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$

<u>Exemple</u> :

Une suspension bactérienne contient 5000 bactéries/litre. On ensemence à partir de cette suspension, 50 **boites de Pétri**, à raison d'1 cm^3 par boite. Si X représente le nombre de colonies par boite, alors la loi de probabilité de X est : $X \to \mathrm{P}(\lambda = 5)$ Boite de Pétri avec colonies bactériennes

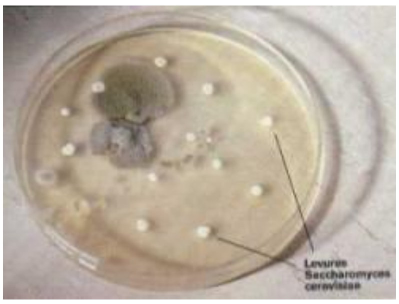
La probabilité qu'il n'y ait **aucune** colonie sur la boite de Pétri est : $P(X = 0) = \dfrac{5^0 e^{-5}}{0!} = 0.0067$

Soit approximativement **0,67 % de chance**.

La probabilité qu'il n'y ait **au moins une** colonie sur la boite de Pétri est :

$P(X > 0) = 1 - P(X = 0) = 1 - 0.0067 = 0.9933$

Soit **99,3 % de chance** d'avoir au moins une colonie bactérienne qui se développe dans la boite de Pétri. (Voir <u>événement contraire</u>)



Comme pour la loi binomiale, il est possible d'utiliser une formule de <u>récurrence</u> pour calculer les valeurs des probabilités successives : $P(X = k) = \dfrac{\lambda}{k} P(X = k - 1)$

2.4.3 Espérance et variance

L'**espérance** d'une variable aléatoire de Poisson est : $E(X)=\lambda$

La **variance** d'une variable de Poisson est : $V(X)=\lambda$

Remarque : il est à noter que dans le cas d'une variable aléatoire de Poisson, **l'espérance et la variance prennent la même valeur**.

Exemples :

(1) Dans le cadre de la culture bactérienne, le nombre moyen de colonies attendu sur la boite de Pétri est : $E(X)=\lambda=5\ \ colonies$

Ainsi si l'on effectue plusieurs cultures bactériennes (plusieurs boites de Pétri) à partir de la même solution initiale, on attend en moyenne **cinq colonies** pour l'ensemble des boites.

En ce qui concerne la variance et l'écart-type, on aura : $V(X)=\lambda=5$ et $\sigma(X)=\sqrt{V(X)}=2.24\ \ colonies$

(2) Sachant que 0,2 % des sujets présentent une allergie lors d'une vaccination, quelles sont les valeurs de l'espérance et de la variance du nombre de sujets atteints lors d'une campagne de vaccination concernant 1000 sujets ?

2.4.4 Stabilité de la loi de Poisson

Si X et Y sont deux variables aléatoires indépendantes suivant des **lois de Poisson** respectivement

$X \rightarrow \mathrm{P}(\lambda)$ et $Y \rightarrow \mathrm{P}(\mu)$ alors $X+Y \rightarrow \mathrm{P}(\lambda+\mu)$

2.5.1 Loi géométrique

Lorsque le nombre de succès n est égal à 1, la loi de la variable aléatoire discrète X porte le nom de loi de Pascal ou **loi géométrique** de paramètre p telle que : $\mathrm{P}(X=k)=pq^{k-1}\ \ ,avec: k\in\mathbb{N}^*$

2.5.2 Espérance et variance

D'où l'**espérance** associée à la loi géométrique est : $E(X)=\dfrac{1}{p}$

la **variance** associée à la loi géométrique est : $V(X)=\dfrac{q}{p^2}$

3 Lois continues

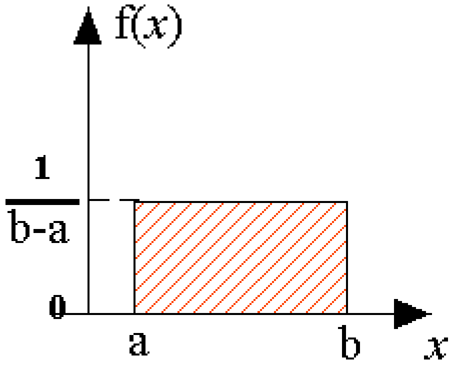
Par définition, les variables aléatoires continues prennent des valeurs continues sur un intervalle donné.

3.1 Loi uniforme

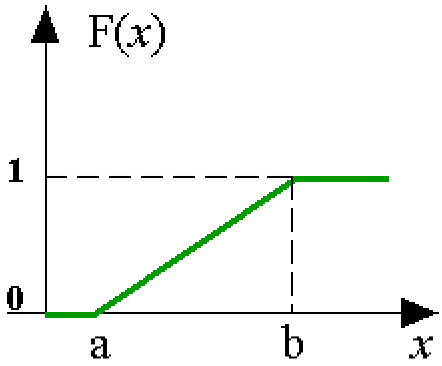
3.1.1 Définition

La loi uniforme est la loi exacte de phénomènes continus uniformément répartis sur un intervalle.

La variable aléatoire X suit une **loi uniforme** sur le segment $[a,b]$ avec $a < b$ si sa densité de probabilité est donnée par : $f(x) = \begin{cases} \dfrac{1}{b-a} & si : x \in [a,b] \\ 0 & si : x \notin [a,b] \end{cases}$



Fonction de densité de probabilité



Fonction de répartition

Quelques commentaires :

(1) La loi uniforme continue étant une loi de probabilité, l'**aire hachurée en rouge** sur la figure ci-dessus vaut **1**. Ceci implique que la valeur prise par $f(x)$ vaut $\dfrac{1}{b-a}$.

(2) La probabilité que $X \in [a',b']$ avec $a' < b'$ et $a',b' \in [a,b]$ vaut :

$$P(a' < X < b') = \int_{a'}^{b'} f(x)dx = \int_{a'}^{b'} \frac{1}{b-a}dx = \frac{b'-a'}{b-a}$$

(3) La **fonction de répartition** associée à la loi uniforme continue est telle que : $F_X(x) = \begin{cases} 0 & si : x < a \\ \dfrac{x-a}{b-a} & si : a \le x \le b \\ 1 & si : x > b \end{cases}$

3.1.2 Espérance et variance

L'**espérance** de la loi uniforme continue vaut : $E(X) = \dfrac{b+a}{2}$

La **variance** de la loi uniforme continue vaut : $V(X) = \dfrac{(b-a)^2}{12}$

3.2 Loi normale ou loi de Laplace-Gauss

3.2.1 Définition

On parle de **loi normale** lorsque l'on a affaire à une variable aléatoire continue dépendant d'un grand nombre de causes indépendantes dont les effets s'additionnent et dont aucune n'est prépondérante (conditions de Borel). Cette loi acquiert sa forme définitive avec **Gauss** (en 1809) et **Laplace** (en 1812). C'est pourquoi elle porte également les noms de : **loi de Laplace, loi de Gauss et loi de Laplace-Gauss.**

Une **variable aléatoire absolument continue** X suit une loi normale de paramètres (m,s) si sa densité de probabilité

$$f:\mathbb{R}\to\mathbb{R}$$

est donnée par :

$$x \mapsto f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Avec $m\in\mathbb{R}$ et $s\in\mathbb{R}^+$ Notation : $X \to \mathrm{N}(m,s)$

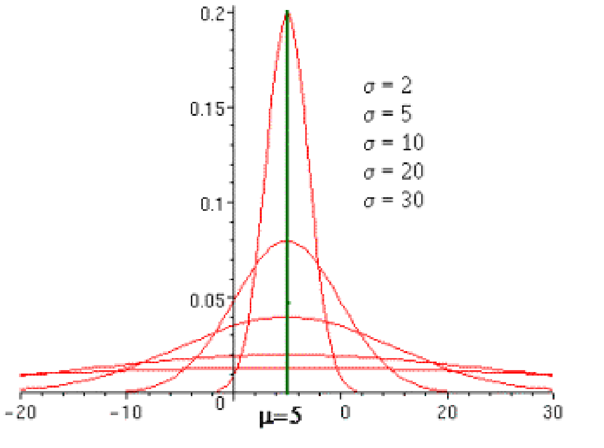
Remarque : On admet que** $\int_{-\infty}^{+\infty} f(x)dx=1$ **dans la mesure où l'intégration analytique est impossible

3.2.2 Etude de la fonction densité de probabilité

La fonction f est **paire** autour d'un axe de symétrie $x=m$ car $f(x+m)=f(x-m)$

d'où $D_E=[m,+\infty[$

x	m	$+\infty$
$f'(x)$	$-$	
$f(x)$	$\frac{1}{\sigma\sqrt{2\pi}}$ ↘	0



Remarque : Le paramètre m représente **l'axe de symétrie** et s **le degré d'aplatissement** de la courbe de la loi normale dont la forme est celle d'une courbe en cloche.

3.2.3 Espérance et variance

L'**espérance** de la loi normale vaut : $E(X)=m$

La **variance** de la loi normale vaut : $V(X)=s^2$

3.2.4 Stabilité de la loi normale

Théorème :

Soient X_1 et X_2 deux variables aléatoires normales indépendantes de paramètres respectifs (m_1,s_1), (m_2,s_2), alors leur somme X_1+X_2 est une variable aléatoire normale de paramètres $\left(m_1+m_2,\sqrt{\sigma_1^2+\sigma_2^2}\right)$

Ce théorème se généralise immédiatement à la **somme de *n* variables aléatoires normales indépendantes.**

3.3 Loi normale réduite

3.3.1 Définition

Une variable aléatoire continue X suit une **loi normale réduite** si sa densité de probabilité est donnée par :

$$f:\mathbb{R}\to\mathbb{R}$$
$$x\mapsto f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

Remarque :

f est bien une loi de probabilité car :

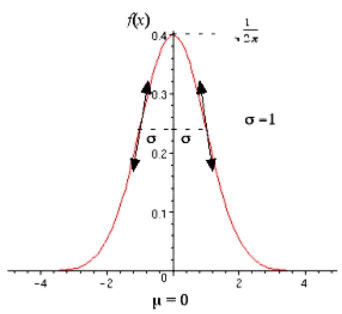
- $\forall x\in\mathbb{R}: f(x)>0$
- f *est intégrable sur* $\mathbb{R}$ *et* $\int_{-\infty}^{+\infty}f(x)dx=1$

3.3.2 Etude de la fonction densité de probabilité

La fonction f est paire car $f(-x)=f(x)$ d'où $D_E=[0,+\infty[$

La dérivé première est $f'(x)=-xf(x)$ avec $f'(x)\leq 0$ pour $x\geq 0$.

x	0 $\qquad$ $+\infty$
$f'(x)$	$-$
$f(x)$	$\frac{1}{\sqrt{2\pi}}$ $\searrow$ 0



Remarque : L'axe de symétrie correspond à l'axe des ordonnées $(x=0)$ et le degré d'aplatissement de la courbe de la loi normale réduite est 1.

3.3.3 Espérance et variance

L'**espérance** d'une loi normale réduite est : $E(X)=0$

La **variance** d'une loi normale réduite est : $V(X)=1$

3.3.4 Relation avec la loi normale

Si X suit une loi normale $\mathrm{N}(m,s)$, alors $Z=\dfrac{X-\mu}{\sigma}$, une <u>variable centrée réduite</u> suit une **la loi normale réduite** $\mathrm{N}(0,1)$

3.3.5 Calcul des probabilités de la loi normale

La **fonction de répartition** de la loi normale réduite permet d'obtenir les probabilités associées à toutes variables aléatoires normales $\mathrm{N}(m,s)$ après transformation en <u>variable centrée réduite.</u>

On appelle **fonction** F_X , la fonction de répartition d'une variable normale réduite X telle que :

$F_X : \mathbb{R} \rightarrow \mathbb{R}$

$$x \mapsto F_X(x) = P(X<x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} f(t)dt$$

Propriétés :

Les **propriétés** associées à la fonction de répartition F_X sont :

(P1) F_X est croissante, continue et dérivable sur $\mathbb{R}$ et vérifie : $\lim_{x\to-\infty} F_X(x)=0$ et $\lim_{x\to+\infty} F_X(x)=1$

(P2) $\forall x \in \mathbb{R} : F_X(x)+F_X(x)=1$

$\forall x \in \mathbb{R} : F_X(x)-F_X(x)=2F_X(x)-1$

Une **application directe** de la fonction F_X est la lecture des probabilités sur la <u>table de la loi normale réduite</u>.

3.4 Lois déduites de la loi normale

3.4.1 Loi du χ^2 de <u>Pearson</u>

Définition

La **loi de Pearson ou loi de χ^2 (Khi deux)** trouve de nombreuses applications dans le cadre de la comparaison de proportions, des tests de conformité d'une distribution observée à une distribution théorique et le test d'indépendance de deux caractères qualitatifs. Ce sont les test du khi-deux.

Soit $X_1, X_2, X_3, ..., X_n$, n variables **normales centrées réduites**, on appelle C^2 la variable aléatoire définie par :

$$C^2 = X_1 + X_2 + X_3 + ... + X_n = \sum_{i=1}^{n} X_i$$

On dit que C^2 suit une **loi de Pearson** à n degrés de liberté (d.d.l.).

Remarque : Si n =1, la variable du C^2 correspond au carré d'une variable normale réduite de loi $\mathrm{N}(0,1)$

Propriétés :

(P_1) Si $X_1, X_2, X_3, ..., X_n$ sont n **variables normales centrées réduites** et s'il existe k relations de dépendance entre ces variables alors $X_1^2 + X_2^2 + X_3^2 + ... + X_n^2$ suit une **loi de Pearson à *n* - *k* degrés de liberté**.

(P_2) Si U suit une loi de Pearson à n d.d.l.,

si V suit une loi de Pearson à m d.d.l.,

Et si U et V sont indépendants alors $U+V$ suit une loi de Pearson à $n+m$ ddl

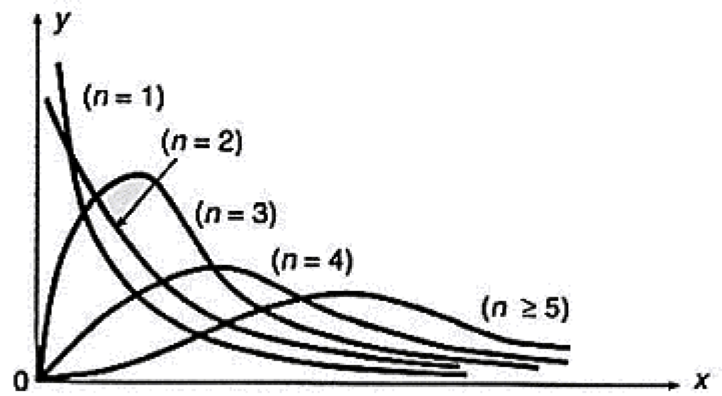
$U-V$ suit une loi de Pearson à $n-m$ ddl (si $n<m$)

Pour $C^2 > 0$, la fonction densité de probabilité est de la forme : $f(\chi^2) = C(n)(\chi^2)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}$

avec : $C(1) = \frac{1}{\sqrt{2\pi}}$

Pour c^2 £ 0 , $f(c^2) = 0$

Pour $n > 1$, on utilise la table du Khi 2



Remarque :**La constante** $C(n)$ **est telle que** $\int_{-\infty}^{+\infty} f(x)dx = 1$ **. La distribution du** C^2 **est dissymétrique et tend à devenir symétrique lorsque *n* augmente en se rapprochant de la <u>distribution normale</u> à laquelle elle peut être assimilée lorsque *n* > 30.**

Espérance et variance

L'**espérance** de la variable du C^2 est : $E(C^2) = n$

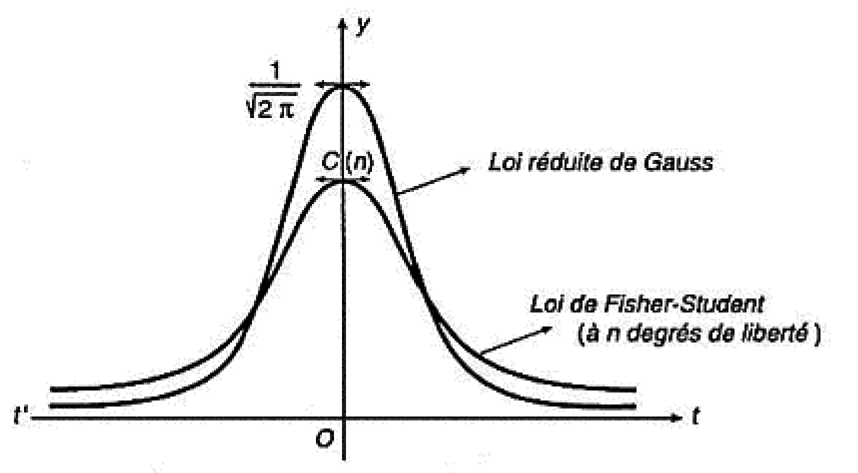
La **variance** de la variable du C^2 est : $V(C^2) = 2n$

3.4.2 Loi de *<u>student</u>*

Définition

La **loi de Student (ou loi de Student-Fisher)** est utilisée lors des tests de comparaison de paramètres comme la moyenne et dans l'estimation de paramètres de la population à partir de données sur une échantillon (Test de Student). Student est le pseudonyme du statisticien anglais William Gosset qui travaillait comme conseiller à la brasserie Guinness et qui publia en 1908 sous ce nom, une étude portant sur cette variable aléatoire.

Soit U une variable aléatoire suivant une **loi normale réduite** $\mathrm{N}(0,1)$ et V une variable aléatoire suivant une **loi de Pearson** à n degrés de liberté C_n^2 , U et V étant **indépendantes**, on dit alors que $T_n = \frac{U}{\sqrt{V/n}}$ suit une **loi de Student** à n degrés de liberté.

La fonction densité de probabilité est de la forme : $f(T) = C(n)\left(1+\frac{T^2}{n}\right)^{-\frac{(n+1)}{2}}$ Calcul des probabilités avec <u>la table de la loi de Student</u>	

Remarque : La constante $C(n)$ est telle que $\int_{-\infty}^{+\infty} f(x)dx = 1$. La distribution du T de Student est symétrique et <u>tend vers une loi normale</u> lorsque n augmente indéfiniment.

Espérance et variance

L'**espérance** de la variable de Student est : $E(T) = 0$

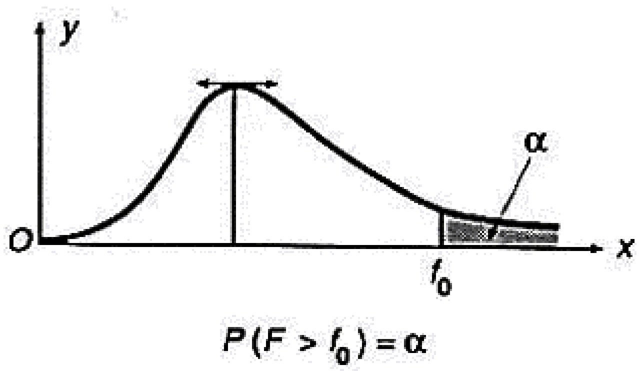
La **variance** de la variable de Student est : $V(T) = \dfrac{n}{n-2}$

3.4.3 Loi de Fisher-Snedecor

La **loi de Fisher-Snedecor** est utilisée pour comparer deux variances observées et sert surtout dans les très nombreux tests d'analyse de variance et de covariance.

Soit U et V deux variables aléatoires indépendantes suivant une **loi de Pearson** respectivement à n et m degrés de liberté.

On dit que $F = \dfrac{U/n}{V/m}$ suit une loi de **Fisher-Snedecor** à $\begin{Bmatrix} n \\ m \end{Bmatrix}$ degrés de liberté.

Pour $F > 0$, la fonction densité de probabilité est de la forme : $f(F) = C_{(n,m)} F^{\frac{n}{2}-1} (m + nF)^{-\frac{(n+m)}{2}}$ Pour $F \le 0$, on a $f(F) = 0$ Utilisation des tables de Fisher-Snedecor pour le calcul des probabilité	

Remarque : **Si $n = 1$, alors on a la relation suivante :** $F_{(1,m)} = \dfrac{U^2}{V/m} = T_m^2$

Espérance et variance

L'**espérance** de la variable de Fisher-Snédecor est : $E(F) = \dfrac{m}{m-2}$ si $m > 2$

La **variance** de la variable de Fisher-Snédecor est : $V(T) = \dfrac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ si $m > 4$

4. Convergence

Dans ce paragraphe, sont traités des éléments de calcul des probabilités dont l'application statistique est nombreuse. La partie fondamentale est **le théorème central limite**. Les éléments présentés permettent de préciser ce que signifie l'ajustement d'une loi de probabilité par une autre loi (notion de convergence) et ainsi de justifier l'approximation d'une distribution observée par une loi théorique. De plus ces éléments permettent de donner des limites d'erreurs possibles dans l'estimation d'un élément d'une population

1. Convergence en loi

Soit une suite de n variables aléatoires $X_1, X_2, X_3, \ldots, X_i, \ldots, X_n$. Cette suite **converge en loi** vers la variable aléatoire X de fonction de répartition F_X quand ***n* augmente indéfiniment**, si la suite des fonctions de répartition $F_{X_1}, F_{X_2}, F_{X_3}, \ldots, F_{X_n}$ tend vers la fonction de répartition F_X pour tout x pour lequel F_X est continue.

2. Le théorème central limite

Appelé également théorème de la limite centrale, il fut établi par Liapounoff et Lindeberg.

On se place dans une situation d'épreuves répétées, caractérisées par une suite $X_1, X_2, X_3, \ldots, X_i, \ldots, X_n$ de n variables aléatoires indépendantes et de même loi (espérance $E(X_i)$ = m et variance $V(X_i) = s^2$). On définit ainsi deux nouvelles variables aléatoires :

La somme $S_n = X_1 + X_2 + \ldots + X_i + \ldots + X_n$

La moyenne $M_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n} = \dfrac{S_n}{n}$

Telles que :

$E(S_n) = nm$ $V(S_n) = ns^2$	$E(M_n) = m$ $V(M_n) = \dfrac{s^2}{n}$

Théorème central limite

Soit la variable aléatoire S_n résultant de la **somme de *n* variables aléatoires Indépendantes et de même loi**, on construit la variable centrée réduite telle que : $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$

Alors pour tout t Î R, la fonction de répartition $F_n(t) = P(Z_n < t)$ est telle que : $F_n(t) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{z^2}{2}} dz$

Quand $n \to +\infty$ c'est à dire N(0,1).

Remarque : On peut calculer Z_n aussi bien à partir de S_n que de M_n car $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{M_n - \mu}{\sigma / \sqrt{n}}$

Une variable aléatoire résultant de la somme de plusieurs v.a. ayant même loi et même paramètres est distribuée suivant une **loi normale réduite** lorsque le nombre d'épreuves ***n* tend vers l'infini**.

Le théorème central limite s'applique **quelque soit la loi de probabilité** suivie par les variables aléatoires discrètes ou continues, pourvu que les épreuves soient **indépendantes, reproductibles et en très grand nombre**.

Grâce au théorème de la limite centrale, on peut voir que des phénomènes dont la variation est engendrée par un nombre important de causes indépendantes sont généralement susceptibles d'être représentés par une loi normale.

A l'aide de la ***convergence en loi*** et du ***théorème central limite***, il est possible de faire l'approximation de certaines lois de probabilités par d'autres

4.3 Convergence vers la loi normale

4.3.1 La loi binomiale

Théorème :

Soit X une variable aléatoire suivant une l**oi binomiale** de paramètres (n,p), alors $P(X = k) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2}$

quand $n \to +\infty$ c'est à dire N(0,1) avec $m = E(X) = np$ et $s^2 = V(X) = npq$

La convergence est d'autant plus rapide que p est voisin de 0,5, distribution symétrique pour la loi binomiale.

Remarque : On considère que l'approximation est valable si on a à la fois $np \geq 5$ et $nq \geq 5$ (voir Rapport entre loi de probabilité).

4.3.2 Loi de Poisson

Théorème :

Soit X une variable aléatoire suivant une **loi de Poisson** de paramètre λ

alors $P\left(X = k\right) \approx \dfrac{1}{\sqrt{\lambda}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\lambda}{\sqrt{\lambda}}\right)^2}$ quand $n \to +\infty$ c'est à dire N(0,1) avec $E\left(X\right) = 1$ et $V(X) = l$

Remarque : On considère qu'on peut faire ces approximations si l ³ 20

4. * Inégalité de Bienaymé-Tchébycheff

L'inégalité de Markov et l'inégalité de Bienaymé-Tchébycheff s'appliquent aussi bien aux variables aléatoires discrètes ou absolument continues. Elles permettent pour une variable aléatoire X d'espérance $E(X)$ = m et de variance σ^2 d'évaluer la probabilité pour que X diffère de la moyenne d'une quantité inférieure à une valeur h.

Le problème est de donner une consistance quantitative à la remarque déjà faite que, plus l'écart-type d'une variable aléatoire est faible, plus sa distribution de probabilité est concentrée autour de son espérance mathématique (voir degré d'aplatissement de la loi normale). Afin de démontrer cette inégalité, nous allons présenter tout d'abord l'inégalité de Markov.

4.1 Inégalité de Markov

Soit X une variable aléatoire admettant une espérance $E(X)$ et une variance $V(X)$, étant donné, un réel $h>0$ **l'inégalité de Markov** donne : $P\left(\left|X\right| \geq h\right) \leq \dfrac{1}{h^2} E\left(X^2\right)$

4.2 Inégalité de Bienaymé-Tchébycheff

Si l'on applique l'inégalité de Markov à la variable aléatoire $X - \overline{X}$, on a $P\left(\left|X - \overline{X}\right| \geq h\right) \leq \dfrac{1}{h^2} E\left(\left[X - \overline{X}\right]^2\right)$

Or $E\left(\left[X - \overline{X}\right]^2\right) = V\left(X\right) = \sigma^2$ et en passant à l'évènement contraire, on a : $P\left(\left|X - \overline{X}\right| \leq h\right) \geq 1 - \dfrac{\sigma^2}{h^2}$

En posant $h = t\sigma$ avec $t > 0$, on obtient l'inégalité suivante $P\left(\left|\dfrac{X - \overline{X}}{\sigma}\right| \leq t\right) \geq 1 - \dfrac{1}{t^2}$ sachant que s > 0

Qui est équivalente à : $P\left(\left|\dfrac{X - \overline{X}}{\sigma}\right| \geq t\right) \leq \dfrac{1}{t^2}$ « $t > 0$ » inégalité de **Bienaymé-Tchébychef**

Remarque : Ces inégalités n'ont d'intérêt que si t est assez grand