VARIABLES ALEATOIRES DISCRETES

A. DAOUI

4 décembre 2012

Définition et caractéristiques

Définition : Une V.A. X sera dite continue si elle peut prendre toutes les valeurs d'un intervalle ou d'une réunion d'intervalles de $\mathbb{R}$. Dans ce cas il existe une fonction $f(x)$ appelée fonction de densité de probabilité telle que pour tout intervalle $[a; b]$ de $\mathbb{R}$, la probabibité que X prenne des valeurs dans $[a; b]$ est l'aire $\int_a^b f(x)dx$.

$$p(a \leq X \leq b) = \int_a^b f(x)dx$$

De plus $f(x)$ doit vérifier les propriétés suivantes :

- $f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

Exemple

Soit X la V.A. de fonction de densité de probabilité

$$\begin{cases} 2x & si \quad 0 < x < 1 \\ 0 & \quad ailleurs \end{cases}$$

Vérifier que f est bien une densité de probabilité et évaluer $p(0.5 \leq X \leq 0.7)$

Fonction de répartition

La fonction de répartition d'une V.A. X est la fonction des probabilités cumulées croissantes ; elle est définie sur $\mathbb{R}$ par :

$$F(x) = p(X < x) = \int_{-\infty}^{x} f(t)dt$$

La fonction F vérifie les propriétés suivantes :

- $0 \leq F(x) \leq 1, \forall x \in \mathbb{R}$
- $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$ et F est croissante

La connaissance de la fonction de répartition permet de donner la loi de probabilité ; En effet :

$$p(a \leq X \leq b) = F(b) - F(a)$$

Calculons la fonction de répartition en reprenant l'exemple précédent.

Espérance et variance

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

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

où

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

En reprenant l'exemple précédent, ona :

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 2x^2dx = \frac{2}{3}$$

$$E(X^2) = \int_0^1 2x^3dx = \frac{1}{2} \Longrightarrow V(X) = \frac{1}{18}$$

On sait qu'une loi discrète est dite uniforme lorsque toutes ses valeurs sont équiprobables. Une V.A. X continue sera dite uniforme lorsque la probabilité que X soit dans un intervalle est directement proportionnelle à la longueur de celui-ci ; c.a.d. la probabilité que X soit dans un intervalle de longueur l est indépendante de la position de l'intervalle dans le domaine $[a, b]$ où X est définie.
Une telle V.A. a pour fonction de densité

$$f(x) = \begin{cases} \frac{1}{b-a} & si \quad x \in [a, b] \\ 0 & \quad ailleurs \end{cases}$$

On note : $X \rightsquigarrow U(a, b)$

On vérifie que

- $F(x) = \begin{cases} 0 & si \quad x < a \\ \frac{x-a}{b-a} & si \quad a \leq x < b \\ 1 & si \quad x \geq b \end{cases}$
- $E(X) = \int_0^1 \frac{x}{b-a} dx = \frac{a+b}{2}$
- $V(X) = \frac{(b-a)^2}{12}$

Soit X le temps (en heures) indiqué par les aiguilles d'une horloge lorsqu'elle s'arrête.
$X \rightsquigarrow U(0, 12)$ car il y a autant de chances que les aiguilles s'arrêtent dans un intervalle entre 0 et 12 que tout autre intervalle de même longueur.
Calculons $p(1 \leq x \leq 4)$, $p(3 \leq x \leq 6)$, $p(3 \leq x \leq 9)$

On a : $F(x) = \begin{cases} 0 & si \quad x < 0 \\ \frac{x}{12} & si \quad 0 \leq x < 12 \\ 1 & si \quad x \geq 12 \end{cases}$

$p(1 \leq x \leq 4) = F(4) - F(1) = \frac{4}{12} - \frac{1}{12} = \frac{1}{4}$

$p(3 \leq x \leq 6) = \frac{1}{4}$

$p(3 \leq x \leq 9) = \frac{1}{2}$

Cette loi est très utilisée pour l'étude des phénomènes d'attente. Elle très liée à la loi de Poisson ; en effet dans le contexte de cette dernière on est amené à considérer le nombre de fois qu'un événment se réalise dans un interrvalle de temps donné alors que l'intervalle de temps qui sépare deux réalisations successives de l'événement est une V.A. continue dont le comportement est décrit par une loi exponentielle. Une V.A. X obéissant à une loi exponentielle de paramètre a $(a > 0)$ a pour fonction de densité :

$$f(x) = \begin{cases} a.e^{-ax} & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases}$$

On note : $X \rightsquigarrow exp(a)$

On vérifie que

- $F(x) = \begin{cases} 0 & si \quad x < 0 \\ 1 - e^{-ax} & si \quad x \geq 0 \end{cases}$
- $E(X) = \frac{1}{a}$ (Intégration par parties)
- $V(X) = \sigma^2 = \frac{1}{a^2}$ (deux fois par parties)

On remarque que $\sigma = m = \frac{1}{a}$ (Similitude avec la loi de Poisson)

Exemple

Dans un petit aéroport, le temps X (exprimé en mn) qui s'écoule entre l'arrivée de deux avions successsifs est distribué selon une loi exponentielle. Sachant que l'intervalle de temps moyen qui sépare 2 avions est de 25 mn, on a les probabilités

$p(X < 25mn) = 0,632$

$p(X > 20mn) = 0,449$

$p(\frac{1}{4}h < X < \frac{1}{2}h) =$

Une V.A. X continue suit une loi normale si sa fonction de densité est de la forme, pour tout $x \in \mathbb{R}$,

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

m et σ sont respectivement la moyenne et l'écart-type de X.

- On remarque que la courbe C_f représentant f est symétrique par rapport à la droite d'équation $x = m$
- Cette loi est la plus importante de toutes les lois ; c'est une loi limite ; la plupart des lois convergent vers elle dès qu'il s'agit d'un qrand nombre d'individus. Cette importance lui a surtout été conférée pa le "Theorem-Cetral-Limite" (Voir chapitre suivant). Elle sert aussi à définir d'autres lois qui ont leur importance en statistiques : Loi Khi-deux, loi de Student ...

Soit $X \rightsquigarrow N(m;\sigma)$
En posant $Z = \frac{x-m}{\sigma}$, on a $Z \rightsquigarrow N(0;1)$ dite loi normale centrée réduite ou loi normale standard dont la fonction de pépartition est répértoriée sur une table et permet de trouver celle de X via le changement de variable

$$p(X < a) = p(Z < \frac{a-m}{\sigma})$$

Utilisation de la table :

Soit $a \geq 0$

- $p(Z < a)$ est directement lue sur la table
- $p(Z < -a) = p(Z > a) = 1 - p(Z < a) = 1 - F(a)$
- $p(-b < Z < -a) = p(a < Z < b)$

Soit $X \rightsquigarrow B(n; p)$ avec n assez grand et p ni proche de 0 ni proche de 1 (idéalement p autour de $\frac{1}{2}$). Alors on peut supposer que

$$X \rightsquigarrow N(m = np; \sigma = \sqrt{np(p-1)})$$

L'approximation est parfaitement valable dès que $n \geq 30$ ou $np \geq 15$ ou $np \geq 5$ et $n(1-p) \geq 5$

Notons qu'on est entrain de remplacer une loi discrète par une loi continue. Pour remédier à ceci, on utilisera la technique suivante :

$$p(X = k) = p(k - \frac{1}{2} < X < k + \frac{1}{2})$$

Soit $X \rightsquigarrow B(1000; 0,4)$
Calculons
$p(X = 430)$
$p(400 < X < 450)$

Théorème de la limite centrale (T.C.L.)

Proposition

Soit $X_1, X_2, ..., X_n$ *des V.A. indépendantes suivant la même loi de probabilité de moyenne* m *et d'écart-type* σ*. Alors les V.A. (somme)* $S_n = X_1 + X_2 + ... + X_n$ *et la V.A. (moyenne)* $\overline{X_n} = \frac{S_n}{n}$ *ont pour caractéristiques :*

$$E(S_n) = n.m \quad V(S_n) = n.\sigma^2$$
$$E(S_n) = m \quad V(S_n) = \frac{\sigma^2}{n}$$

Théorème

(T.C.L.)

Soit $X_1, X_2, ..., X_n$ *des V.A. indépendantes suivant la même loi de probabilité de moyenne* m *et d'écart-type* σ*. Alors pour* n *assez grand* $(n \geq 30)$*, on peut supposer que :*

- $S_n \rightsquigarrow N(nm; \sqrt{n}.\sigma)$
- $\overline{X_n} \rightsquigarrow N(m; \frac{\sigma}{\sqrt{n}})$

Exemple

Durant une matinée, 400 étudiants se sont présentés au centre-copie. Chacun d'entre eux a payé une somme X_i. On suppose que les X_i des V.A. indépendantes de même loi de probabilité de moyenne $m = 10Dh$ et d'écart-type $\sigma = 5Dh$. Donner la probabilité poir que la recette totale vdu centre-copie dépasse $4200Dh$.

Loi de Student

Une V.A. suit une loi de Student à n degrés de liberté si elle est obtenue par le rapport :

$$T = \frac{X}{\sqrt{X_1 + X_2 + ... + X_n}},$$

où $X, X_1, X_2, ..., X_n$ sont des V. A ; indépendantes suivant la loi normale centrée réduite $N(0; 1)$.
On note $T \rightsquigarrow St_n$

- Notons que St_n présente les mêmes propriétés de symétrie que $N(0;1)$
- $E(St_n) = 0$ et $V(St_n) = \frac{n}{n-2}$
- St_n tend vers $N(0;1)$. On estime que St_n se comporte comme $N(0;1)$ dès que $n \geq 30$.
- La loi de Student joue un rôle important dans les sondages et les tests statistiques. Elle prend la place de la loi normale de variance inconnue dans la cas de petits échantillons.
- Il exite une table dannant les, pour différentes probabilités α, la borne t tel que $p(St_n < t) = \alpha$.

Exemple

- Calculer $p(St_{14} < 1,34)$ et $p(St_{30} < 0,53)$ (comparer avec $p(Z < 0,53)$)
- Déterminer t tel que $p(St_{15} < t) = 0,8$ puis $p(St_{20} > t) = 0,1$

Exemple