Université Paris 1
Magistère d'Economie - 1ère année

COURS DE STATISTIQUE

DEFINITIONS ET PROPRIETES DES PRINCIPALES LOIS UNIDIMENSIONNELLES

Convention : Si la variable aléatoire (v.a.) X suit la loi $\mathcal{L}$, on notera $X \sim \mathcal{L}$.

Lois discrètes

1. Loi UNIFORME : $\mathcal{U}$
Elle est définie sur des entiers $\{1, ..., n\}$ ou des valeurs $\{a_1, ..., a_n\}$ avec la même probabilité pour toutes les valeurs
$$\mathbb{P}(X = i) = \frac{1}{n} \ , \ i = 1, ..., n$$
ou
$$\mathbb{P}(X = a_i) = \frac{1}{n} \ , \ i = 1, ..., n.$$
Espérance et variance :
Si $X(\Omega) = \{1, ..., n\}$,
$$\mathbb{E}(X) = \frac{n+1}{2} \ , \ \mathbb{V}(X) = \frac{n^2 - 1}{12}$$
Si $X(\Omega) = \{a_1, ..., a_n\}$,
$$\mathbb{E}(X) = \frac{1}{n}\sum_{i=1}^{n} a_n \ , \ \mathbb{V}(X) = \frac{1}{n}\sum_{i=1}^{n} a_i^2 - \mathbb{E}(X)^2$$

2. Loi de BERNOULLI : $\mathcal{B}(1, p)$
On utilise cette loi lorsque les résultats possibles d'une épreuve aléatoire sont réduits à deux :
Oui-Non ; Vrai-Faux ; Succès-Echec
On parle dans ce cas d'expérience de Bernoulli. Une v.a. X suit une loi $\mathcal{B}(1, p)$ si elle prend deux valeurs 0 et 1 avec les probabilités suivantes :
$$\mathbb{P}(X = 1) = p \ , \ \mathbb{P}(X = 0) = 1 - p = q$$
La loi de Bernoulli peut aussi s'écrire sous la forme suivante :
$$\mathbb{P}(X = k) = p^k(1-p)^{1-k} \ , \ k \in \{0, 1\}$$
Espérance et variance :
$$\mathbb{E}(X) = p \ , \ \mathbb{V}(X) = p(1-p)$$
Fonction génératrice :
$$G(s) = \mathbb{E}\left(s^X\right) = q + ps$$

3. Loi BINOMIALE : $\mathcal{B}(n,p)$
Une v.a. suivant cette loi peut se définir comme le nombre de "succès" à l'issue de n expériences de Bernoulli indépendantes et de même loi. Le support d'une telle variable est $X(\Omega) = \{0, 1, ..., n\}$ avec la loi de probabilité :

$$P(X = k) = C_n^k p^k (1-p)^{n-k} \ , \ k \in \{0, 1, ..., n\}$$

Construction
Si n v.a. indépendantes $X_1, X_2, ..., X_n$ suivent la même loi $\mathcal{B}(1,p)$, alors la variable $X = \sum_{i=1}^{n} X_i$ suit une loi Binomiale $\mathcal{B}(n,p)$.
Espérance et variance :

$$\mathbb{E}(X) = np \ , \ \mathbb{V}(X) = np(1-p)$$

Fonction génératrice :

$$G(s) = \mathbb{E}\left(s^X\right) = (q + ps)^n$$

Addition
Si deux v.a. indépendantes X_1 et X_2 suivent respectivement les lois $\mathcal{B}(n_1,p)$ et $\mathcal{B}(n_2,p)$, alors la variable aléatoire $X = X_1 + X_2$ suit une loi $\mathcal{B}(n_1 + n_2, p)$.
Approximation (Convergence en loi)
Lorsque n est grand :

   (a) Si $n > 30$, si np et $np(1-p)$ sont voisins et si $np \leq 15$, on peut alors approximer la loi $\mathcal{B}(n,p)$ par une loi de Poisson $\mathcal{P}(\lambda)$, où $\lambda = np$.

   (b) Si $np > 15$ et $n(1-p) > 15$ (ou si $np(1-p) > 5$), on peut alors approximer la loi $\mathcal{B}(n,p)$ par une loi normale $\mathcal{N}\left(m, \sigma^2\right)$, où $m = np$ et $\sigma^2 = np(1-p)$.

4. Loi GEOMETRIQUE : $\mathcal{G}(p)$
Une v.a. suivant cette loi peut se définir comme le nombre d'essais jusqu'au premier succès pour des épreuves de Bernoulli indépendantes de paramètre p. Le support d'une telle variable est $X(\Omega) = \mathbb{N}^\star$ avec la loi de probabilité :

$$P(X = k) = p(1-p)^{k-1} \ , \ k \in \mathbb{N}^\star$$

Espérance et variance :

$$\mathbb{E}(X) = \frac{1}{p} \ , \ \mathbb{V}(X) = \frac{1-p}{p^2}$$

Fonction génératrice :

$$G(s) = \mathbb{E}\left(s^X\right) = \frac{sp}{1-qs} \ , \ q = 1-p$$

NB : On définit parfois la loi Géométrique comme le nombre d'échecs avant le premier succès. Dans ce cas, elle prend ses valeurs dans $\mathbb{N}$,
$P(X = k) = p(1-p)^k \ , \ k \in \mathbb{N}$, $\mathbb{E}(X) = \frac{1}{p} - 1$ et $\mathbb{V}(X) = \frac{1-p}{p^2}$. Quand on ne précise rien, on considère la 1ère définition.

5. Loi de POISSON : $\mathcal{P}(\lambda)$
Une v.a. X suit une loi de Poisson si elle prend ses valeurs dans $\mathbb{N}$ avec les probabilités :

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{x!} \ , \ k \in \mathbb{N}$$

Le paramètre λ est toujours considéré positif, $\lambda > 0$.
Espérance et variance :

$$\mathbb{E}(X) = \mathbb{V}(X) = \lambda$$

Fonction génératrice :

$$G(s) = \mathbb{E}\left(s^X\right) = e^{\lambda(s-1)}$$

Addition
Si deux v.a. indépendantes X_1 et X_2 suivent respectivement les lois $\mathcal{P}(\lambda_1)$ et $\mathcal{P}(\lambda_2)$, alors la variable aléatoire $X = X_1 + X_2$ suit une loi $\mathcal{P}(\lambda_1 + \lambda_2)$.

Lois continues

6. Loi UNIFORME : $\mathcal{U}(a,b)$
Une v.a. X suit une loi uniforme sur l'intervalle $[a,b]$ si X est une v.a. continue de densité f constante :

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

Espérance et variance :

$$\mathbb{E}(X) = \frac{a+b}{2} \ , \ \mathbb{V}(X) = \frac{(b-a)^2}{12}$$

7. Loi EXPONENTIELLE : $\mathcal{E}(\lambda)$
Une v.a. X suit une loi exponentielle $\mathcal{E}(\lambda)$ si elle est continue et de densité :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , \quad x > 0 \\ 0 & , \quad x \leq 0 \end{cases}$$

Espérance et variance :

$$\mathbb{E}(X) = \frac{1}{\lambda} \ , \ \mathbb{V}(X) = \frac{1}{\lambda^2}$$

8. Loi de LAPLACE-GAUSS ou loi NORMALE : $\mathcal{N}\left(m, \sigma^2\right)$
Une v.a. X suit une loi exponentielle $\mathcal{N}\left(m, \sigma^2\right)$ si elle est continue et de densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \quad , \quad x \in \mathbb{R}$$

Espérance et variance :

$$\mathbb{E}(X) = m \ , \ \mathbb{V}(X) = \sigma^2$$

Lecture de table :
Une v.a. X de loi $\mathcal{N}\left(m, \sigma^2\right)$ peut se mettre sous la forme suivante :

$$X \sim \mathcal{N}\left(m, \sigma^2\right) \ \Leftrightarrow \ U = \frac{X-m}{\sigma} \sim \mathcal{N}(0,1) \ ,$$

loi centrée réduite (ou standard) qui est tabulée ; on a donc

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(U \leq \frac{X-m}{\sigma}\right)$$

Addition :
Si deux v.a. **indépendantes** X_1 et X_2 suivent respectivement des lois normales $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$, alors la v.a. $X_1 + X_2$ suit une loi normale $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Combinaison linéaire :
On démontre que si $X_1, X_2, ..., X_n$ suivent des lois normales **indépendantes** d'espérances $m_1, m_2, ..., m_n$ et de variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$, alors pour tous $\lambda_i \in \mathbb{R}$, la v.a. $\sum_{i=1}^n \lambda_i X_i$ suit une loi normale d'espérance $\sum_{i=1}^n \lambda_i m_i$ et de variance $\sum_{i=1}^n \lambda_i^2 \sigma_i^2$.

Lois de v.a. liées à la loi Normale

Certaines lois peuvent être construites à partir de la loi Normale. Il n'est pas utile de donner ici une définition complète de ces lois par leur densité.

9. Loi du CHI-DEUX : χ_n^2 (Chi-deux à n degrés de liberté)
Soit $U_1, U_2, ..., U_n$ n v.a. indépendantes et de même loi $\mathcal{N}(0,1)$. Alors,
$$Z = \sum_{i=1}^n U_i^2 \ \sim \ \chi_n^2$$
Espérance et variance :
$$\mathbb{E}(\chi_n^2) = n \ , \ \mathbb{V}(\chi_n^2) = 2n$$
Addition :
Si deux v.a. **indépendantes** Y_1 et Y_2 suivent respectivement des lois $\chi_{n_1}^2$ et $\chi_{n_2}^2$, alors la v.a. $Y_1 + Y_2$ suit une loi $\chi_{n_1+n_2}^2$.

Approximation :
Lorsque n est grand ($n > 30$), la loi de la v.a. $\sqrt{2\chi_n^2} - \sqrt{2n-1}$ peut être approximée par une loi Normale centrée et réduite $\mathcal{N}(0,1)$.

10. Loi de STUDENT : $\mathcal{T}_n$ (Student à n degrés de liberté)
Soit X et Y deux v.a. **indépendantes** telles que X suit une loi $\mathcal{N}(0,1)$ et Y suit une loi χ_n^2. Alors,
$$Z = \frac{X}{\sqrt{\frac{Y}{n}}} \ \sim \ \mathcal{T}_n$$
avec $\mathbb{E}(Z) = 0$.

Approximation :
Lorsque n est grand ($n > 60$), la loi $\mathcal{T}_n$ peut être approximée par une loi Normale centrée et réduite $\mathcal{N}(0,1)$.

11. Loi de FISHER-SNEDECOR : $\mathcal{F}(n_1, n_2)$ à n_1 et n_2 degrés de liberté.
Soit X_1 et X_2 deux v.a. **indépendantes** de lois respectives $\chi_{n_1}^2$ et $\chi_{n_2}^2$. Alors, la v.a.
$$Z = \frac{X_1/n_1}{X_2/n_2}$$
suit une loi $\mathcal{F}(n_1, n_2)$ à n_1 et n_2 degrés de liberté.

Lois de v.a. liées à la loi Exponentielle

12. Loi GAMMA : $\gamma(\alpha, \lambda)$

Une v.a. X suit une loi Gamma de paramètres α et λ, pour $\alpha > 0$ et $\lambda > 0$, si X est une v.a. continue de densité f :

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & , \quad x > 0 \\ 0 & , \quad x \leq 0] \end{cases} ,$$

où $\Gamma(\alpha)$ est une constante de normalisation, définie par :

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx \ .$$

(Cette intégrale est convergente pour $\alpha > 0$ et si $\alpha \in \mathbb{N}^\star$, alors $\Gamma(\alpha) = (\alpha - 1)!$).

Espérance et variance :

$$\mathbb{E}(X) = \frac{\alpha}{\lambda} \ , \ \mathbb{V}(X) = \frac{\alpha}{\lambda^2}$$

Addition :
Si deux v.a. **indépendantes** X_1 et X_2 suivent respectivement des lois $\gamma(\alpha_1, \lambda)$ et $\gamma(\alpha_2, \lambda)$, alors la v.a. $X_1 + X_2$ suit une loi $\gamma(\alpha_1 + \alpha_2, \lambda)$.

Relations entre les différentes lois :
La loi $\mathcal{E}(\lambda)$ n'est autre que la loi $\gamma(1, \lambda)$.
La somme de n v.a. Exponentielles **indépendantes** de paramètre λ suit une loi $\gamma(n, \lambda)$.
De même, la loi $\chi^2(1)$ n'est autre que la loi $\gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ et, par conséquent, la loi $\chi^2(n)$ se confond avec la loi $\gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

Propriétés de multiplication par une constante :
Si une v.a. X suit une loi $\gamma(\alpha, \lambda)$ et si μ est un scalaire positif, alors la v.a. μX suit une loi $\gamma\left(\alpha, \frac{\lambda}{\mu}\right)$.

Conséquence :
Si une v.a. X suit une loi $\gamma(n, \lambda)$, alors la v.a. $2\lambda X$ suit une loi $\gamma\left(n, \frac{1}{2}\right)$, c'est-à-dire une loi $\chi^2(2n)$. On peut donc se servir des tables de la loi du χ^2 pour les sommes de v.a. exponentielles **indépendantes**.

PROPRIETES DES N-ECHANTILLONS

Définition d'un n-échantillon :
$(X_1, X_2, ..., X_n)$ forment un n-échantillon de X si les v.a. $X_1, X_2, ..., X_n$ sont indépendantes et suivent toutes la loi de X.

Définition d'une statistique :
Une statistique est une fonction "mesurable" des v.a. de l'échantillon ; c'est une variable aléatoire.
Une fonction mesurable est telle que $f^{-1}(I) \in \mathcal{A}$, pour tout intervalle I, c'est-à-dire telle qu'on sache calculer $\mathbb{P}\left(f^{-1}(I)\right)$ pour tout intervalle I.

Cas gaussien

Un échantillon

Soit $(X_1, X_2, ..., X_n)$ un n-échantillon de X, où la v.a. $X \sim \mathcal{N}(m, \sigma^2)$. Alors, on a les résultats suivants :

1. $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$$
2. $$\frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i - m)^2 \sim \chi^2_n$$
3. $$\frac{1}{\sigma^2}\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \sim \chi^2_{n-1}$$
4. $$\frac{\overline{X} - m}{\frac{S'}{\sqrt{n}}} \sim \mathcal{T}_n \ ,$$
   où $S'^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - m)^2$.
5. $$\frac{\overline{X} - m}{\frac{S}{\sqrt{n}}} \sim \mathcal{T}_{n-1} \ ,$$
   où $S^2 = \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$.
6. $\overline{X}$ et S^2 sont des v.a. indépendantes.

Deux échantillons

Soit $(X_1, X_2, ..., X_{n_1})$ et $(Y_1, Y_2, ..., Y_{n_2})$ deux échantillons indépendants de lois respectives $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$. Alors, on a les résultats suivants :

7. $$\frac{\frac{\sum_{i=1}^{n_1}(X_i-m_1)^2}{n_1\sigma_1^2}}{\frac{\sum_{j=1}^{n_2}(Y_j-m_2)^2}{n_2\sigma_2^2}} \sim \mathcal{F}(n_1, n_2)$$

8. $$\frac{\frac{\sum_{i=1}^{n_1}\left(X_i-\overline{X}\right)^2}{(n_1-1)\sigma_1^2}}{\frac{\sum_{j=1}^{n_2}\left(Y_j-\overline{Y}\right)^2}{(n_2-1)\sigma_2^2}} \sim \mathcal{F}(n_1-1, n_2-1)$$

9. $$\frac{\left(\overline{X}-\overline{Y}\right)-(m_1-m_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0,1)$$

10. Si $\sigma_1 = \sigma_2$, alors

$$\frac{\left(\overline{X}-\overline{Y}\right)-(m_1-m_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{\sum_{i=1}^{n_1}\left(X_i-\overline{X}\right)^2+\sum_{j=1}^{n_2}\left(Y_j-\overline{Y}\right)^2}{n_1+n_2-2}}} \sim \mathcal{T}_{n_1+n_2-2}$$