Variable aléatoire réelle

Discrète

Distribution de probabilité $p_i = P(X = x_i)$

$$\sum_{i=1}^{n\ ou\ \infty} p_i = 1$$

Fonction de répartition

$$F(u) = P(X \leq u) = \sum_{x_k \leq u} P(X = x_k)$$

Espérance

$$E(X) = \mu = \sum_{\omega \in \Omega} X(\omega).P\big(X(\omega)\big)$$

Variance

$$Var(X) = \sigma^2 = \sum_{\omega \in \Omega} (X(\omega) - \mu)^2.P\big(X(\omega)\big) = \left[\sum_{\omega \in \Omega} X(\omega)^2.P\big(X(\omega)\big)\right] - \mu^2$$

Continue

$f(x)$ = **densité de probabilité** ssi :

- f est continue sur $\mathbb{R}$
- $f(x) \geq 0, \forall x \in \mathbb{R}$
- $\int_{-\infty}^{+\infty} f(x)\, dx = 1$

Le couple de v.a.r $Z = (X, Y)$ est dit constitué de deux v.a.r **indépendantes** ssi :

$$f_Z(x, y) = f_X(x) \times f_Y(y)$$

Fonction de répartition

$$F(u) = P(X \leq u) = \int_{-\infty}^{u} f(x)\, dx$$

F est continue et dérivable ($F' = f$) sur $\mathbb{R}$

$$F(x) \in [0,1], \forall x \in \mathbb{R}$$

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to +\infty} F(x) = 1$$

$P(a \leq X \leq b) = F(b) - F(a)$
$P(X = a) = 0$
$P(X > a) = 1 - P(X \leq a)$

Espérance

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

Variance

$$Var(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 . f(x)\, dx = \int_{-\infty}^{+\infty} x^2 . f(x) dx - \mu^2$$

Propriétés

Si l’espérance est nulle, la v.a.r est dite **centrée**
$E(aX + b) = aE(X) + b$

La variance est une mesure de dispersion, elle est positive ou nulle
La racine carrée de la variance, notée σ, est appelée **écart-type**
Si la variance est égale à 1, la v.a.r est dite **réduite**

$Var(X + a) = Var(X)$
$Var(aX + b) = a^2Var(X)$

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

Lois de probabilités discrètes

Loi uniforme $X \sim \mathcal{U}(1, n)$

$$P(X = k) = \frac{1}{k}$$

$$E(X) = \frac{n+1}{2}$$

$$Var(X) = \frac{n^2 - 1}{12}$$

Loi de Bernouilli (succès / échec) $X \sim \mathcal{B}(1, p)$

$$P(X = x) = \begin{cases} p \quad si\ x = 1 \\ q = 1 - p \quad si\ x = 0 \end{cases}$$

$$E(X) = p$$

$$Var(X) = p.q$$

Loi Binomiale $X \sim \mathcal{B}(n, p)$

La v.a.r est le nombre total de succès n expériences indépendantes de même probabilité p
Cette loi est aussi appelée loi des épreuves répétées

$$P(X = k) = \binom{n}{k} . p^k . q^{n-k}$$

$$E(X) = n.p$$

$$Var(X) = n.p.q$$

$$P(X = k + 1) = P(X = k) . \frac{(n - k).p}{(k + 1).q}$$

Si X et Y sont **indépendantes** $X \sim \mathcal{B}(n_1, p)$ et $Y \sim \mathcal{B}(n_2, p)$

Alors $X + Y \sim \mathcal{B}(n_1 + n_2, p)$

Si $X_1, X_2, \dots, X_n$ sont **indépendantes** et $X_i \sim \mathcal{B}(1, p)$

Alors $Y = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n, p)$

Application du théorème central limite

Si pour $Y \sim \mathcal{B}(n, p)$ $n \geq 30, np \geq 5$ et $nq \geq 5$

Alors $Y \sim \mathcal{N}\left(np, \sqrt{np(1-p)}\right)$

$$P(Y = k) \approx P\left(k - \frac{1}{2} \leq X \leq k + \frac{1}{2}\right) = \pi\left(\frac{k + \frac{1}{2} - np}{\sqrt{npq}}\right) - \pi\left(\frac{k - \frac{1}{2} - np}{\sqrt{npq}}\right)$$

correction de continuité

Loi de Poisson $X \sim \mathcal{P}(\lambda)$

Loi associée aux succès des évènements rares

$$P(X = k) = \frac{\lambda^k . e^{-\lambda}}{k!}$$

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

$$P(X = k + 1) = P(X = k) . \frac{\lambda}{k + 1}$$

En pratique, on utilise souvent l'**approximation** suivante :
(meilleur approximation que celle de la loi normale pour « p petit »)

Si $X \sim \mathcal{B}(n, p)$ avec $n > 20$ et $p < 0{,}5$

Alors $X \sim \mathcal{P}(\lambda = np)$

Application du théorème central limite

Si pour $Y \sim \mathcal{P}(\lambda)$ $\lambda \geq 20$

Alors $Y \sim \mathcal{N}(\lambda, \sqrt{\lambda})$

$$P(X = k) \approx P\left(k - \frac{1}{2} \leq Y \leq k + \frac{1}{2}\right) = \pi\left(\frac{k + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \pi\left(\frac{k - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right)$$

correction de continuité

Lois de probabilités continues

Loi uniforme (ou loi rectangulaire) $X \sim \mathcal{U}(a,b)$

$b > a$

$$f(x) = \begin{cases} \dfrac{1}{b-a} \ si\ a \leq x \leq b \\ 0\ sinon \end{cases}$$

$$F(x) = \begin{cases} 0\ \ si\ x \leq a \\ \dfrac{x-a}{b-a}\ si\ a \leq x \leq b \\ 1\ si\ x > b \end{cases}$$

$$E(X) = \frac{a+b}{2}$$

$$Var\ (X) = \frac{(b-a)^2}{12}$$

Loi exponentielle $X \sim \mathcal{E}(\theta)$

$\theta > 0$

Utilisé dans le cas où la v.a.r. X décrit des évènements aléatoires évoluant dans le temps (ex : durées de vies)

$$f(x) = \begin{cases} \theta e^{-\theta x}\ si\ x \geq 0 \\ 0\ sinon \end{cases}$$

$$P(X \leq x) = F(x) = 1 - e^{-\theta x} \quad si\ x > 0$$

$$E(x) = \frac{1}{\theta}$$

$$Var\ (X) = \frac{1}{\theta^2}$$

Loi Normale (ou de Gauss) $\quad X \sim \mathcal{N}(\mu, \sigma)$

$\sigma > 0$

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

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

$$X \sim \mathcal{N}(\mu,\sigma) \Leftrightarrow U = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

U est dite **loi Normale centrée réduite**

$$\pi(u) = P(U \leq u) = F(u)$$

$$P(U \geq u) = P(U \leq -u) = \pi(-u) = 1 - P(U \leq u) = 1 - \pi(u)$$

$$P(-u \leq U \leq u) = 1 - 2\pi(-u) = 2\pi(u) - 1$$

Si X et Y sont **indépendantes** : $\quad X \sim \mathcal{N}(\mu_1, \sigma_1)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2)$

Alors : $\quad X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

Mais : $\quad X - Y \sim \mathcal{N}(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

Soit une suite de $X_1, X_2, \dots, X_n$ de n v.a.r. **indépendantes et de même loi Normale** $\quad X_i \sim \mathcal{N}(\mu, \sigma)$

Alors la loi de probabilité de la v.a.r moyenne :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Vérifie : $\quad \overline{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$

Loi du chi-deux à n degrés de liberté

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2$$

Avec $X_i \sim \mathcal{N}(0,1)$

$$E(\chi_n^2) = n$$

$$Var(\chi_n^2) = 2n$$

Pour 2 lois du chi-deux **indépendantes** à n et p ddl, on a :

$$\chi_n^2 + \chi_p^2 = \chi_{n+p}^2$$

Remarque : $X \sim \mathcal{N}(0,1) \Rightarrow X^2 \sim \chi_1^2$

Donc $P(X^2 \leq u) = P\left(-\sqrt{u} \leq X \leq \sqrt{u}\right) = \pi\left(\sqrt{u}\right) - \pi\left(-\sqrt{u}\right)$

Loi de Student à n degré de liberté $\qquad T \sim \mathcal{T}(n)$

$$T = \frac{U}{\sqrt{\frac{V}{n}}}$$

Avec $\quad U \sim \mathcal{N}(0,1) \quad$ et $\quad V \sim \chi_n^2$

$$E(\mathcal{T}(n)) = 0$$

Application du théorème central limite

Si pour $\quad X \sim \mathcal{T}(n) \qquad n > 30$

Alors $\quad X \sim \mathcal{N}(0,1)$

Loi de Fischer à n et p degrés de liberté $\qquad Z \sim \mathcal{F}(n,p)$

$$Z = \frac{X/n}{Y/p}$$

Avec $\quad X \sim \chi_n^2 \quad$ et $\quad Y \sim \chi_p^2$

$$E(\mathcal{F}(n,p)) = \frac{p}{p-2}$$

Théorème de la limite centrale

Soit une suite de $X_1, X_2, \ldots, X_n$ de n v.a.r. **indépendantes et de même loi**, dont on note μ l'espérance commune et σ^2 la variance commune.

Alors la loi de probabilité de la v.a.r. somme :

$$S_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

Tend vers (converge) : $\mathcal{N}(n\mu, \sqrt{n}\sigma)$ lorsque n augmente (« indéfiniment »)

Posons :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Le théorème central limite permet d'affirmer qu'asymptotiquement :

$$\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$$

Cette loi est utilisée pratiquement dès $n > 30$, si l'on connaît μ et σ, mais **la loi des X_i n'est pas forcément une loi Normale**

Estimation ponctuelle et par intervalle

Un « bon estimateur » T d'un paramètre θ doit être le plus proche possible de sa vraie valeur.
Il faut donc fournir des critères de qualité entre T et θ

- **Définition du biais**

Un estimateur T d'un paramètre θ est dit sans biais si :

$$E(T) = \theta$$

Si ce n'est pas le cas, l'estimateur est dit biaisé

- **Définition de la convergence**

Un estimateur T_n (dépendant de la taille n de l'échantillon) d'un paramètre θ est dit convergent s'il est sans biais et si :

$$\lim_{n \to +\infty} E[(T_n - \theta)^2] = 0$$

- **Définition de l'écart quadratique moyen**

L'erreur quadratique moyenne d'un estimateur est :

$$EQM(T) = E[(T - \theta)^2]$$

Si T est sans biais :

$$EQM(T) = var(T)$$

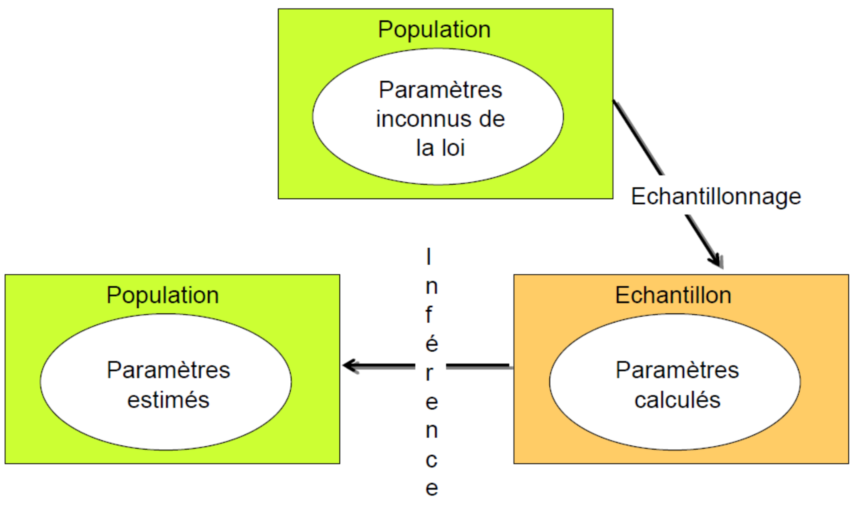
- **Précision d'un estimateur** $= \sqrt{var(T)}$

- **Définition d'un intervalle de confiance**

On appelle intervalle de confiance d'un paramètre θ un intervalle $[\theta_i, \theta_s]$ tel que :

$$P(\theta \in [\theta_i, \theta_s]) = 1 - \alpha$$

α : risque d'erreur
$1 - \alpha$: niveau de confiance ou seuil



Loi Binomiale

Estimateur de p

Il est sans biais et convergent

$$\hat{p}_n = \frac{X}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n} \qquad var(\hat{p}_n) = \frac{p.(1-p)}{n}$$

Intervalle de confiance avec $n < 30$

$$P(X < n_0) = \sum_{i=0}^{n_0-1} \binom{n}{i} \left(p_{inf}\right)^i (1 - p_{inf})^{n-i} = 1 - \frac{\alpha}{2}$$

$$P(X \leq n_0) = \sum_{i=0}^{n_0} \binom{n}{i} \left(p_{sup}\right)^i (1 - p_{sup})^{n-i} = \frac{\alpha}{2}$$

Intervalle de confiance avec $n \geq 30$

$$IC_{1-\alpha} = \left[p_0 \pm c_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}} \right]$$

Pour $\alpha = 5\%$, $c_{\alpha/2} = 1{,}96$

Loi de Poisson

Estimateur de λ

Il est sans biais et convergent

$$\hat{\lambda}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Loi Normale

Estimateur de la moyenne μ

Il est sans biais et convergent

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad var(\overline{X}_n) = \frac{\sigma^2}{n}$$

Estimateur de la variance σ^2 pour un échantillon (μ connue)

$$s^2 = \hat{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Estimateur de la variance σ^2 pour la population (μ inconnue)

Il est sans biais

$$S^2 = \hat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Formule simplifiée pour le calcul de la variance :

$$S^2 = \hat{S}_n^2 = \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} X_i^2\right) - n.\overline{X}_i^2\right]$$

En pratique :

$$S^2 = \frac{n}{n-1}.s^2$$

Pour n grand : $S^2 \approx s^2$

Intervalle de confiance de l'espérance ($n < 30$ et σ inconnu)

$$IC_{1-\alpha} = \left[\overline{x} \pm t_{n-1,\alpha/2} . \frac{S}{\sqrt{n}}\right]$$

n : taille de l'échantillon

$\overline{x}$: moyenne de l'échantillon

$t_{n-1,\alpha/2}$: table de Student à (n-1) ddl

$S = \sqrt{S^2} = \sqrt{\frac{n}{n-1} . s^2}$

Intervalle de confiance de l'espérance ($n \geq 30$ et σ connu)

$$IC_{1-\alpha} = \left[\overline{x} \pm c_{\alpha/2} . \frac{S}{\sqrt{n}}\right]$$

$S^2 = s^2$ car n grand

Intervalle de confiance de la variance

$$P\left(\chi^2_{n-1,\alpha/2} \leq a\right) = \frac{\alpha}{2} \qquad et \qquad P\left(\chi^2_{n-1,1-\alpha/2} \leq b\right) = 1 - \frac{\alpha}{2}$$

χ^2_{n-1} : table du chi-deux à (n-1) ddl

$$IC_{1-\alpha} = \left[(n-1) . \frac{S^2}{b} ; (n-1) . \frac{S^2}{a}\right]$$

$S^2 = \frac{n}{n-1} . s^2$