Loi de probabilité - discrète et continue.
Fiche Résumé

I. Loi binomiale

Une variable aléatoire X suit une loi binomiale de paramètres n et p,lorsque X compte les succès lors de n répétitions indépendantes d'une épreuve de Bernoulli de paramètre p. On note $X \hookrightarrow \mathscr{B}(n, p)$.

Dans ce cas, $E(X) = np$ et $\sigma(X) = \sqrt{np(1-p)}$, enfin $p(X = k) = \binom{k}{n} p^k (1-p)^{n-k}$

II. Loi Uniforme

Une variable aléatoire X suit une loi uniforme sur l'intervalle $[a, b]$ si et seulement si elle admet pour densité la fonction constante f définie sur $[a, b]$ par

$$f(x) = \frac{1}{b-a}$$

En particulier si $I = [c, d]$ est un intervalle contenu dans $[a, b]$ on a :

$$p(c \leq X \leq d) = \int_c^d \frac{1}{b-a} dt = \frac{d-c}{b-a}$$

On note $X \hookrightarrow U(a, b)$; on a $E(X) = \dfrac{a+b}{2}$.

III. Loi Exponentielle

Une variable aléatoire T suit une loi exponentielle de paramètre $\lambda > 0$ sur l'intervalle $[0; +\infty[$ si et seulement si elle admet pour densité la fonction f définie sur $[0, +\infty[$ par

$$f(x) = \lambda e^{-\lambda x}$$

En particulier si $a > 0$ on a :

$$p(T \leq a) = \int_0^a \lambda e^{-\lambda t} dt = 1 - e^{-\lambda a} \qquad \text{et} \quad p(T \geq a) = 1 - p(T < a) = e^{-\lambda a}$$

Enfin $E(T) = \lim\limits_{x \to +\infty} \int_0^x \lambda t e^{-\lambda t} dt = \dfrac{1}{\lambda}$.

Une loi exponentielle modélise la durée de vie d'un phénomène sans mémoire, ou sans vieillissement, ou sans usure : la probabilité que le phénomène dure au moins s + t heures sachant qu'il a déjà duré t heures sera la même que la probabilité de durer s heures à partir de sa mise en fonction initiale. En d'autres termes, le fait que le phénomène ait duré pendant t heures ne change rien à son espérance de vie à partir du temps t. Mathématiques cela se traduit par :

$$p_{T \geq t}(T \geq t + s) = p(T \geq s)$$

IV. Loi Normale

En théorie des probabilités et en statistique, la loi normale est l'une des lois de probabilité les plus adaptées pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires.

Une variable aléatoire X suit une loi normale centrée réduite si et seulement si elle admet pour densité la fonction f définie sur $]-\infty, +\infty[$ par

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

La courbe de cette densité est appelée courbe de Gauss ou courbe en cloche, entre autres. C'est la représentation la plus connue de cette loi.

On a $E(X) = 0$ et $\sigma(X) = 1$ et on note $X \hookrightarrow \mathcal{N}(0;1)$.

Quelque soit l'intervalle $I = [a, b]$ de $\mathbb{R}$ on a :

$$p(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

probabilité que l'on détermine à l'aide de la calculatrice.

De plus, comme $p(X \leq 0) = p(X \geq 0) = \frac{1}{2}$ (symétrie de la courbe), si $\alpha > 0$ alors

$$p(X < \alpha) = p(X \leq 0) + p(0 \leq X \leq \alpha) = \frac{1}{2} + p(0 \leq X \leq \alpha)$$

Théorème 1.

Soit X une variable aléatoire suivant la loi normale centrée réduite $\mathcal{N}(0;1)$. Pour tout réel $\alpha \in]0;1[$, il existe un unique réel positif u_α tel que :

$$p(-u_\alpha \leq X \leq u_\alpha) = 1 - \alpha$$

On retiendra les deux valeurs suivantes :

— Si $\alpha = 0,05$ alors $p(-1,96 \leq X \leq 1,96) \simeq 0,95$

— Si $\alpha = 0,01$ alors $p(-2,58 \leq X \leq 2,58) \simeq 0,99$

Le lien entre les histogrammes de la loi binomiale et la courbe de Gauss est donné par le théorème suivant :

Théorème 2. (Théorème De Moivre-Laplace)

Si la variable $X_n \hookrightarrow \mathcal{B}(n, p)$ alors la variable $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ converge en loi vers une loi normale centrée et réduite $\mathcal{N}(0,1)$ i.e $\lim\limits_{n\to+\infty} p(a < Z_n < b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

Dans la pratique, on estime que quand X suit une loi $\mathcal{B}(n, p)$ avec $n \geq 30$, $np \geq 5$ et $n(1-p) \geqslant 5$ on peut approximer la variable $Z_n = \frac{X - np}{\sqrt{np(1-p)}}$ par une variable Z de loi $\mathcal{N}(0,1)$ puis se ramener à la loi binomiale de X.

Si $X \hookrightarrow \mathcal{B}(50, 0.2)$ alors

$$p(15 \leq X \leq 30) = p\left(\frac{15-10}{\sqrt{400}} \leq \frac{X-10}{\sqrt{400}} \leq \frac{30-10}{\sqrt{400}}\right) = p\left(\frac{1}{4} \leq \frac{X-10}{\sqrt{400}} \leq 1\right) \simeq \int_{0.25}^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \simeq 0,2$$

V. Intervalle de fluctuation Asymptotique et intervalle de confiance

V.1. Intervalle de fluctuation asymptotique

Exposé du problème : On considère une expérience aléatoire à deux issues possibles (succès ou échec) dont on **connaît** la probabilité de succès p.

Si on réalise cette expérience n fois, en notant X_n la variable aléatoire qui comptabilise le nombre de succès, la proportion de succès est donné par $F_n = \dfrac{X_n}{n}$.

Théorème 3.

L'intervalle $I_n = \left[p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ contient F_n avec une probabilité qui se rapproche de $1-\alpha$ lorsque n augmente.

— On dit alors que α est le risque et I_n est un **intervalle de fluctuation asymptotique** au seuil $1-\alpha$.

— Pour un risque de $\alpha = 5\%$ l'intervalle de fluctuation I_n vaut :

$$\left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

— Pour un risque de $\alpha = 1\%$ l'intervalle de fluctuation I_n vaut :

$$\left[p - 2,58\frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 2,58\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

— Dans la pratique on utilise un intervalle de fluctuation asymptotique au seuil $1-\alpha$ lorsque $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

Exemple : L'entreprise A annonce que le pourcentage de moteurs défectueux dans la production est égal à 1%. Afin de vérifier cette affirmation 800 moteurs sont prélevés au hasard. On constate que 15 moteurs sont détectés défectueux. Le résultat de ce test remet-il en question l'annonce de l'entreprise A ? Justifier. On pourra s'aider d'un intervalle de fluctuation.

Solution : Pour une proportion p et un échantillon de taille n, l'intervalle de fluctuation asymptotique au seuil de 95 % est :

$$\left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

sous les trois conditions : $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

L'échantillon de l'enquête est de taille $n = 800$ et l'entreprise annonce que le pourcentage de moteurs défectueux est égal à 1 % donc $p = 0,01$.

Regardons si les trois conditions sont vérifiées :
$n = 800 \geq 30$, $np = 800 \times 0,01 = 8 \geq 5$ et $n(1-p) = 800 \times 0,99 = 792 \geq 5$.

L'intervalle est : $I = \left[0,01 - 1,96\dfrac{\sqrt{0,01(1-0,01)}}{\sqrt{800}} ; 0,01 + 1,96\dfrac{\sqrt{0,01(1-0,01)}}{\sqrt{800}}\right] \approx [0,003 ; 0,017]$.

On constate que 15 moteurs sont détectés défectueux sur 800, ce qui fait une proportion de $\dfrac{15}{800} = 0,01875$; or $0,01875 \notin I$ donc le résultat de ce test remet en question l'annonce de l'entreprise A.

V.2. Intervalle de confiance

Exposé du problème : On considère une expérience aléatoire à deux issues possibles (succès ou échec) dont on **ne connaît pas** la probabilité de succès p.

Si on réalise cette expérience n fois, en notant X_n la variable aléatoire qui comptabilise le nombre de succès, la proportion de succès est donné par $F_n = \dfrac{X_n}{n}$.

Théorème 4.

L'intervalle $I_n = \left[F_n - \dfrac{1}{\sqrt{n}}; F_n + \dfrac{1}{\sqrt{n}}\right]$ contient p avec une probabilité qui se rapproche de 95% lorsque n augmente.

— L'intervalle I_n est appelé intervalle de confiance pour p au niveau asymptotique 95%.

— Dans la pratique on utilise cet intervalle de confiance lorsque $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

Exemple : Lors d'un scrutin électoral, on souhaite connaître la proportion p de français votant pour le candidat « A ». Un institut de sondage mène une enquête auprès de 1000 personnes tirées au hasard (compte tenu de la grandeur de la population, on assimile ce tirage à un tirage avec remise). Le résultat estime qu'une proportion $f = 49\%$ d'entre elles voteront pour le candidat « A ».

Peut-on affirmer d'après l'enquête que le candidat « A » n'aura pas la majorité des votes, c'est-à-dire que $p < 0,5$?

Solution : La condition $n = 1000 \geq 30$ est satisfaite, de plus $np \geq 5 \iff p \geq 0,005$ et $n(1-p) \geq 5 \iff p \leq 0,995$. Il semble raisonnable d'admettre que ces deux dernières conditions sont vérifiées.

L'intervalle de confiance au niveau 95% pour l'estimation de p est donc :

$$\left[0,49 - \frac{1}{\sqrt{1000}}; 0,49 + \frac{1}{\sqrt{1000}}\right] = [0,458; 0,522]$$

Ainsi nous pouvons affirmer que dans 95% des cas p est compris entre 0,458 et 0,522. Il ne paraît donc pas raisonnable d'affirmer suite à cette enquête que le candidat « A » n'obtiendra pas la majorité des votes.

VI. La loi normale $\mathcal{N}(\mu;\sigma^2)$

Soient un réel μ et un réel strictement positif σ. On dit qu'une variable aléatoire X suit la loi normale $\mathcal{N}(\mu;\sigma^2)$ si et seulement si la variable aléatoire $\frac{X-\mu}{\sigma} \hookrightarrow \mathcal{N}(0;1)$. Il suit alors que $E(X) = \mu$ et $\sigma(X) = \sigma$.

Propriété 1.

Soit une variable aléatoire X qui suit la loi normale $\mathcal{N}(\mu,\sigma^2)$. On a :

— $P(\mu-\sigma \leq X \leq \mu+\sigma) \simeq 0.683$

— $P(\mu-2\sigma \leq X \leq \mu+2\sigma) \simeq 0.954$

— $P(\mu-3\sigma \leq X \leq \mu+3\sigma) \simeq 0.997$