Séance 1

Probabilités et statistiques

Image/SCIA EPITA

Plan

Généralités

- étude d'expériences aléatoires,
- expérience aléatoire : expérience dont on ne peut prévoir l'issue à l'avance mais dont on connaît toutes les issues possibles,
- aléa vient du latin *alea* qui est un jeu de dés.

Situation élémentaire

1. n issues $\omega_1,\ldots,\omega_n$,
2. univers $\Omega = \{\omega_1, \cdots, \omega_n\}$,
3. probabilités (d'occurence) associées $p_1,\ldots,p_n$,
4. loi de probabilité : donnée des p_i,
5. les probabilités p_i sont positives et vérifient : $p_1 + \cdots + p_n = 1$,
6. évènement : sous-ensemble de Ω,
7. probabilité d'un évènement : somme des probabilité des issues qui le réalisent.

Exemples d'expériences aléatoires

- lancer d'une pièce,
- expérience de Bernoulli à deux issues : « Succès » et « Echec » aussi appelée *tirage de Bernoulli*,
- lancer d'un dé,
- choix d'une personne dans une population.

- *équiprobabilité* : toutes les issues ayant la même probabilité,
- *exercice* : proposer une situation qui n'est pas équiprobable,
- $A \cap B$: ensemble des issues qui réalisent simultanément A et B,
- $A \cup B$: ensemble des issues qui réalisent au moins un des deux événements A et B,
- A et B sont *indépendants* si $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$.

Conditionnement

- Soient A et B deux événements (supposons $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(B) \neq 0$),
- $\mathbb{P}_A(B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$,
- **Formule de Bayes** : $\mathbb{P}_B(A) = \dfrac{\mathbb{P}_A(B) \times \mathbb{P}(A)}{\mathbb{P}_A(B) \times \mathbb{P}(A) + \mathbb{P}_{\bar{A}}(B) \times P(\bar{A})}$.

Variables aléatoires discrètes

1. variable aléatoire X : fonction définie sur Ω et à valeurs dans $\mathbb{R}$,
2. X peut prendre les valeurs $x_1, \ldots, x_n$,
3. Ω sera "oublié" et on se concentrera sur les probabilités $p_i := \mathbb{P}(\{\omega \in \Omega | X(\omega) = x_i\}) := \mathbb{P}(X = x_i)$,
4. loi d'une variable aléatoire : donnée des réels $\mathbb{P}(X = x_i)$.
5. *exercice* : modéliser le gain à un jeu de Pile ou Face à l'aide d'une variable aléatoire (gain de 100 euros si "Pile" et perte de 80 euros si "Face"),

1. **Attention, la définition des réels p_i a changé !**
2. Espérance : $E(X) = \displaystyle\sum_{i=1}^{n} p_i x_i$,
3. espérance : "valeur moyenne" aussi notée $\bar{x}$,
4. Variance : $V(X) = \sum_{i=1}^{n} p_i (x_i - \bar{x})^2$
5. variance : moyenne pondérée des écarts quadratiques à la moyenne,
6. variance : paramètre de dispersion,
7. écart-type : $\sigma(X) = \sqrt{V(X)}$.

1. Par définition : $V(X) = E((X - E(X))^2)$,
2. On obtient, après un calcul, $V(X) = E(X^2) - E(X)^2$,
3. $E(X^k) = \sum_{i=1}^{n} x_i^k p_i$,
4. $m_k := E(X^k)$ moment d'ordre k,
5. moments comme indicateurs de dispersion,
6. $\mu_k := E((X - E(X))^k)$,
7. μ_k : moment centré d'ordre k,
8. μ_3 : *skewness* (coefficient d'asymétrie) en finance,
9. μ_4 : *kurtosis* (coefficient d'aplatissement) en finance.

1. Deux variables aléatoires X et Y sont *indépendantes* si, pour tous entiers j et k,
2. $\mathbb{P}(\{X = j\} \cap \{Y = k\}) = \mathbb{P}(\{X = j\}) \times \mathbb{P}(\{Y = k\})$.

Loi de Bernoulli

1. variable aléatoire X pouvant prendre les valeurs 0 et 1,
2. probabilité de prendre la valeur 1 notée p,
3. par conséquent : $\mathbb{P}(X = 0) = 1 - p$.
4. $E(X) = p$ et $V(X) = p(1 - p)$,
5. loi notée $\mathcal{B}(p)$.

Loi binomiale de paramètres n et p

1. nombre de succès après n répétitions indépendantes d'une expérience de Bernoulli,
2. autre définition : somme de n variables aléatoires indépendantes suivant une loi $\mathcal{B}(p)$,
3. variable aléatoire X pouvant prendre les valeurs entières comprises entre 0 et n,
4. $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ pour $k \in \{0, \cdots, n\}$.
5. loi notée $\mathcal{B}(n, p)$
6. $\binom{n}{k}$: coefficient binomial,
7. $E(X) = np$ et $V(X) = np(1-p)$.

Loi binomiale négative de paramètres n et p

1. aussi appelée *loi de Pascal*,
2. nombre d'échecs nécessaires avant d'obtenir n succès,
3. variable aléatoire X pouvant prendre toutes les valeurs entières,
4. $\mathbb{P}(X = k) = \binom{k+n-1}{n-1} p^n (1-p)^k$ pour $k \in \{0, \cdots, n\}$.
5. $\binom{n}{k}$: coefficient binomial,
6. $E(X) = n\dfrac{1-p}{p}$ et $V(X) = n\dfrac{1-p}{p^2}$.

Loi géométrique de paramètre p

1. nombre d'essais avant le premier succès dans une répétition de tirages indépendants de Bernoulli,
2. p : probabilité de "Succès",
3. X peut prendre toutes les valeurs entières hormis zéro,
4. $\mathbb{P}(X = k) = pq^{k-1}$ où $q = 1 - p$,
5. $E(X) = \dfrac{1}{p}$ et $V(X) = \dfrac{q}{p^2}$,
6. loi notée $\mathcal{G}(p)$.

Loi de Poisson de paramètre λ

1. X peut prendre toutes les valeurs entières,
2. λ paramètre strictement positif,
3. $\mathbb{P}(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$,
4. $E(X) = \lambda$ et $V(X) = \lambda$,
5. loi notée $\mathcal{P}(\lambda)$.

Cadre

- X définie sur l'univers Ω et à valeurs dans $\mathbb{R}$ (ou dans un intervalle I),
- $\mathbb{P}(X \in [a; b]) = \int_a^b f(x)dx$,
- fonction f appelée la *densité* de la variable aléatoire X,
- Pour un réel x donné : $\mathbb{P}(X = x) = 0$.

Densité de probabilité

1. 2 conditions à connaître :
2. $f(x) \geq 0$, pour tout réel $x \in I$,
3. $\int_I f(x)dx = 1$ (L'intervalle I peut être $\mathbb{R}$) .

Fonction de répartition

- Soit X une variable aléatoire,
- $F_X(x) := \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)dt$
- Fonction de survie : $R_X(x) := \mathbb{P}(X > x) = 1 - F_X(x)$.

Espérance

- formule analogue au cas discret,
- si $\int_I |x| f(x) dx < +\infty$, la variable aléatoire X est dite *intégrable*,
- $E(X) = \int_I x f(x) dx$ existe et définit l'espérance de la v.a. X,
- propriétés de linéarité valables.

Variance

- Si $\int_I x^2 |f(x)| dx < +\infty$, la variable aléatoire X est dite de carré *intégrable*,
- $V(X) = \int_I (x - E(X))^2 f(x) dx$ est bien définie,
- $V(X) = E(X^2) - E(X)^2$ (théorème de Koenig-Huyghens),
- $V(aX) = a^2 V(X)$.

Loi uniforme sur l'intervalle $[a; b]$

1. $f(x) = \dfrac{1}{b-a}$ pour $x \in [a; b]$ et $f(x) = 0$ pour $x \notin [a; b]$,
2. $E(X) = \dfrac{a+b}{2}$ et $V(X) = \dfrac{(b-a)^2}{12}$,
3. *Exercice* : démontrer ce résultat puis calculer la fonction de répartition associée.
4. loi notée $\mathcal{U}([a; b])$.

Loi exponentielle de paramètre $\lambda > 0$

1. $f(x) = \lambda e^{-\lambda x}$ pour $x \geq 0$ et $f(x) = 0$ pour $x < 0$,
2. $E(X) = \dfrac{1}{\lambda}$ et $V(X) = \dfrac{1}{\lambda^2}$,
3. $F(x) = 1 - e^{-\lambda x}$ pour $x \geq 0$ et $F(x) = 0$ sinon,
4. $R(x) = e^{-\lambda x}$ pour $x \geq 0$ et $R(x) = 1$ sinon.

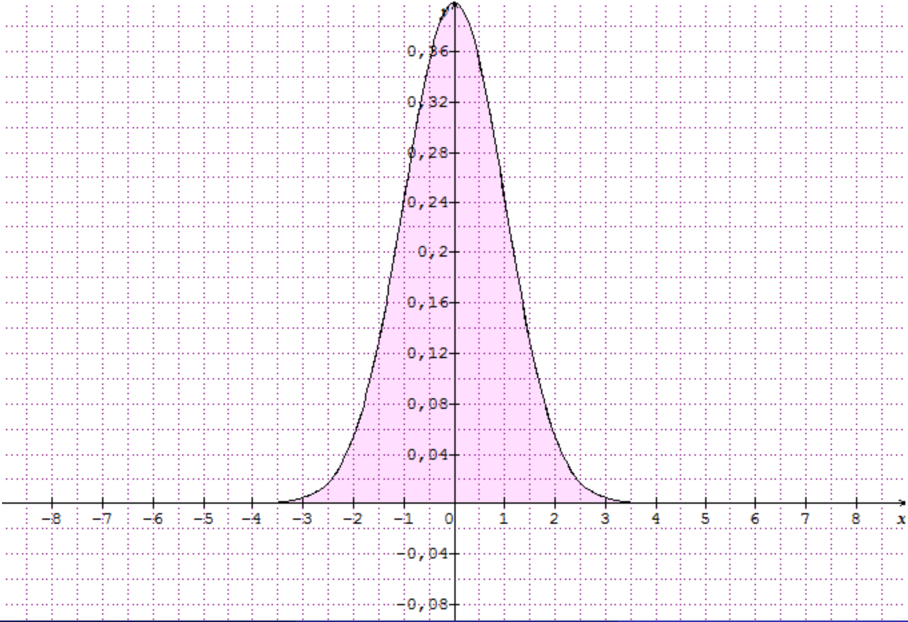
Loi exponentielle

1. loi notée $\mathcal{E}(\lambda)$,
2. durée de vie d'un phénomène sans mémoire,
3. $\forall s \geq 0, \forall t \geq 0, \mathbb{P}_{\{T>t\}}(T > s+t) = \mathbb{P}(T > s)$,
4. modélise aussi l'arrivée des clients dans une file d'attente.

Loi normale centrée réduite

1. $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ pour $x \in \mathbb{R}$,
2. $E(X) = 0$ et $V(X) = 1$,
3. loi notée $\mathcal{N}(0; 1)$.

Loi normale centrée réduite



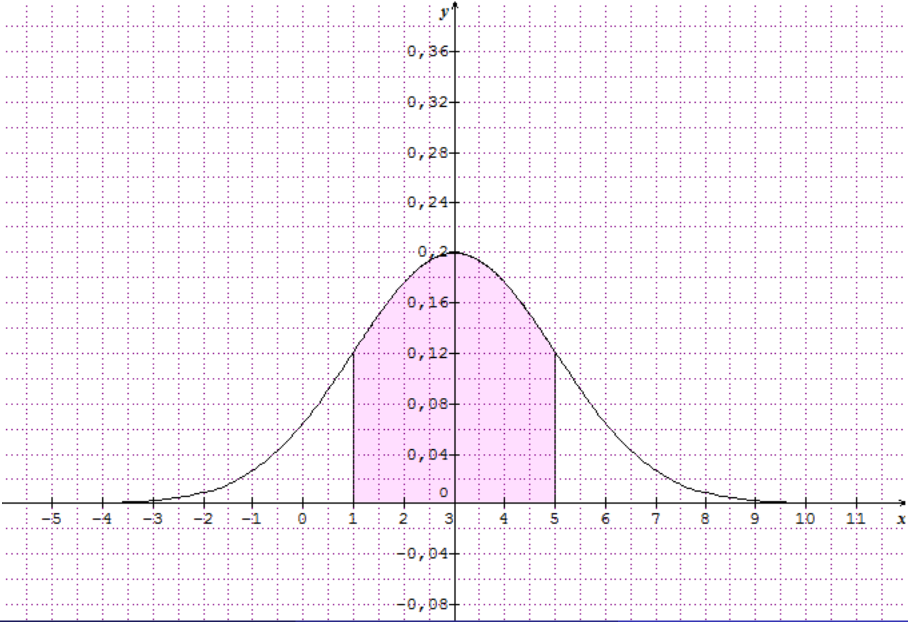
Loi normale centrée réduite

1. $\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = 0,5$,
2. $\mathbb{P}(X \leq -a) = \mathbb{P}(X \geq a)$,
3. $\mathbb{P}(-1,96 \leq X \leq 1,96) \approx 0,95$ et $\mathbb{P}(-2,58 \leq X \leq 2,58) \approx 0,99$
4. loi notée $\mathcal{N}(0;1)$.

Loi normale de paramètres μ et σ

1. loi notée $\mathcal{N}(\mu, \sigma^2)$,
2. X suit une loi $\mathcal{N}(\mu, \sigma^2)$ si $Y = \dfrac{X - \mu}{\sigma}$ suit une loi normale centrée réduite.
3. densité : $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ pour $x \in \mathbb{R}$,
4. $E(X) = \mu$ et $V(X) = \sigma^2$.

Loi normale de paramètres μ et σ



Loi normale de paramètres μ et σ

1. $\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0,683$,
2. $\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0,954$,
3. $\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0,997$.

Soit X une v.a. continue et g une fonction (mesurable).
Sous la condition $\int_I |g(x)|f(x)dx < +\infty$, on a :

$$\boxed{E(g(X)) = \int_I g(x)f(x)dx}$$

- Enoncé similaire pour le cas discret.
- théorème déjà utilisé pour calculer $E(X^2)$ (cf supra).

Formule du changement de variable

Théorème

Sous les conditions suivantes :

1. $f(x) > 0$ *pour tout réel* $x \in I$ *où* I *est un intervalle ouvert,*
2. g *est bijective de* I *sur* $g(I)$*,*
3. g *est dérivable sur l'intervalle* I*,*

la variable aléatoire $Y = g(X)$ *admet pour densité :*

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f(g^{-1}(y))$$