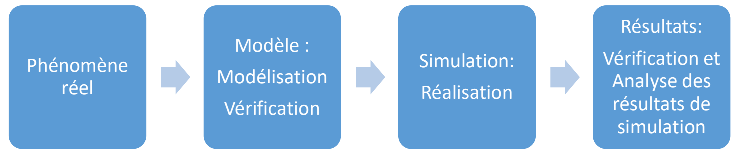
Simulation à événements discrets

4A STI

Etapes de simulation



Simulation

- Un programme qui crée un environnement virtuel afin d'étudier un problème réel
- Situations d'utilisation
  - Expérimentation réel est difficile à réaliser
    - Guerre, système de banque
  - Moins cher
    - RTL simulation pour IC design
    - Planning d'autoroute
  - ……

Rappel : Modèle statistique dans la simulation

Un monde : aléatoire vs statique

Modélisation:

- Choisir une distribution paramétrique
- Estimer le(s) paramètre(s)
- Tester la distribution obtenue

Rappel:

- Quelques distributions importantes

Variables aléatoires discrètes

Une variable aléatoire X est une variable aléatoire si le nombre de valeurs possibles de X est finis ou dénombrable.

Exemples: considérons des clients dans une boutique

- X: le nombre de clients arrivés chaque jour dans la boutique
- R_X: les valeurs possibles de X (espèce de X)

  {0,1,2, …}
- $P(X = x_i) = p(x_i)$: la probabilité de $X = x_i$
- $p(x_i), i = 0,1,2 \ldots$ doivent satisfaire :
  - $p(x_i) \geq 0 \ \ \forall i$
  - $\sum_{i=0}^{\infty} p_i = 1$
- La collection des paires $[x_i, p(x_i)]$ pour , $i = 0,1,2 \ldots$ est appelée la probabilité de distribution de X et $p(x_i)$ est appelé la fonction de masse de X.

Variables aléatoires continues

Exemple: la durée de vie d'une machine X est une variable aléatoire continue avec la densité

$$f(x) = \begin{cases} \dfrac{e^{-\frac{x}{2}}}{2} & x \geq 0 \\ 0 & sinon \end{cases}$$

- X suit une loi exponentielle avec une moyenne de 2 ans
- La probabilité de la durée de vie de cette machine est entre 2 et 3 ans est donnée par

$$P(2 < X < 3) = \frac{1}{2}\int_2^3 e^{-x/2} dx = 0{,}14$$

Variables aléatoires continues

Une variable aléatoire X est une variable aléatoire si son espèce R_X est un intervalle ou une collection des intervalles

La probabilité de $X \in [a, b]$ est donnée par:

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

$f(x)$ est la fonction de densité de X et doit satisfaire

- $f(x) \geq 0 \ \ \forall x \in R_X$
- $\int_{R_X} f(x)dx = 1$
- $f(x) = 0$ si $x \notin R_X$

Propriétés :

- $P(X = x_0) = 0$ parce que $\int_{x_0}^{x_0} f(x)dx = 0$
- $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$

Fonction de distribution

- Une fonction $F(x)$ est appelé une fonction de distribution si $F(x) = P(X \leq x)$
  - Si X est discrète, alors $F(x) = \sum_{x_i \leq x} p(x_i)$
  - Si X est continue, alors $F(x) = \int_{-\infty}^{x} f(t)dt$
- Propriétés
  - F est une fonction non-décroissante: si $a < b$, alors $F(a) \leq F(b)$
  - $\lim_{x\to\infty} F(x) = 1$
  - $\lim_{x\to-\infty} F(x) = 0$
- La fonction de distribution porte toutes les informations sur la distribution de X.

Fonction de distribution

Exemple : la durée de vie d'une machine a une fonction de distribution

$$F(x) = \frac{1}{2}\int_0^x e^{-t/2}dt = 1 - e^{-\frac{x}{2}}$$

- La probabilité que la machine fonctionne moins de 2 ans $P(0 \leq X < 2) = F(2) - F(0) = F(2) = 1 - e^{-1} = 0{,}632$
- La probabilité que la machine fonctionne entre 2 et 3 ans $P(2 < X < 3) = F(3) - F(2) = \left(1 - e^{-\frac{3}{2}}\right) - (1 - e^{-1}) = 0{,}145$

Variance

La variance de X : $Var(X)$

- $Var(X) = E\big(X - E(X)\big)^2$
- $Var(X) = E(X^2) - E(X)^2$
- Une mesure de la dispersion ou la variation des valeurs possibles de X autour de l'espérance

Espérance

L'espérance de X : $E[X]$

- Si X est discrète, $E[X] = \sum_i x_i p(x_i)$
- Si X est continue, $E[X] = \int_{-\infty}^{\infty} xf(x)dx$
- La moyenne ou le moment de degré 1 de X
- Une mesure de la tendance centrale et la gravité centrale

Espérance & Variance

Exemple: la durée de vie de la machine X

- Espérance

$$E(X) = \frac{1}{2}\int_0^\infty xe^{-x/2}dx = 2$$

- Variance

$$E(X^2) = \frac{1}{2}\int_0^\infty x^2e^{-x/2}dx = 8$$

$$Var(X) = E(X^2) - E(X)^2 = 4$$

$$\sigma = \sqrt{Var(X)} = 2$$

Distributions discrètes

Si X est une variable aléatoire discrète, on donne la loi de probabilité de X en donnant une suite (pk) où $p_k = P(X = k)$.

- Exemples
  - Loi de Bernoulli
  - Loi binomiale
  - Loi binomiale négative
  - Loi géométrique
  - Loi de Poisson

Loi de Bernoulli

On dit que X suit la loi de Bernoulli de paramètre p si

$$P(X = 1) = p$$
$$P(X = 0) = 1 - p$$

On a alors $\forall x \in \{0,1\}$

$$P(X = x) = p^x(1-p)^{1-x}$$

- Espérance

$$E(X) = p$$

- Variance

$$Var(X) = p(1-p)$$

- Exemple: pour un événement A,

$$1_A \sim B(P(A))$$

Expériences de Bernoulli

- Considérons une expérience de n essais, chaque essai peut être un succès ou une défaillance

$$X_i = \begin{cases} 1 & si\ le\ i-\text{ème}\ essai\ est\ un\ succès \\ 0 & si\ le\ i-\text{ème}\ essai\ est\ une\ déaillance \end{cases}$$

La distribution de Bernoulli d'un essai X_i, $i = 1,2, \ldots$ est

$$p_i(x_i) = p(x_i) = \begin{cases} p, & x_i = 1 \\ 1 - p = q, & x_i = 0 \end{cases}$$

- Processus de Bernoulli

L'ensemble de n expériences de Bernoulli indépendantes

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)p_2(x_2) \ldots p_n(x_n)$$

Loi binomiale

Considérons $X_1, \ldots, X_n$ n variables suivant la loi de Bernoulli de paramètre p indépendantes.

La somme de $X_1, \ldots, X_n$ ou le nombre de succès

$$S = X_1 + \cdots + X_n$$

suit une loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$.

Pour $k \in \{0, \ldots, n\}$, on a

$$p(k) = P(S = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Loi binomiale

Considérons $X_1, \dots, X_n$ n variables suivant la loi de Bernoulli de paramètre p indépendantes.

La somme de $X_1, \dots, X_n$ ou le nombre de succès

$$S = X_1 + \cdots + X_n$$

suit une loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$.

- Espérance
$$E(S) = E(X_1) + \cdots + E(X_n) = np$$
- Variance
$$Var(S) = Var(X_1) + \cdots + Var(X_n) = np(1-p)$$

Loi binomiale négative

X: Le nombre d'essais de Bernoulli jusqu'au $k$ème succès

- Distribution
$$p(x) = \binom{x-1}{k-1} q^{x-k} p^k, \qquad x = k, k+1, \dots$$
- Espérance
$$E(X) = \frac{k}{p}$$
- Variance
$$Var(X) = \frac{k(1-p)}{p^2}$$

Loi géométrique

X : le nombre d'essais pour arriver le premier succès

Distribution
$$p(x) = q^{x-1} p, \qquad x = 1,2, \dots$$

- Espérance
$$E(X) = \frac{1}{p}$$
- Variance
$$Var(X) = \frac{1-p}{p^2}$$

Loi de Poisson

On dit que X suit la loi de Poisson de paramètre λ si

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \qquad x = 0,1, \dots.$$

$$F(x) = \sum_{i=0}^{x} \frac{e^{-\lambda} \lambda^i}{i!} \qquad x = 0,1, \dots.$$

- Espérance
$$E(X) = \lambda$$
- Variance
$$Var(X) = \lambda$$

Loi de Poisson

Exemple : les tâches arrivées dans une machine avec un taux constant

Distributions continues

- Exemples
  - Loi uniforme
  - Loi normale
  - Loi exponentielle
  - Loi de Weibull
  - Loi lognormale

Loi uniforme

- Une variable aléatoire X suit une loi uniforme en intervalle (a,b), $U(a,b)$ si sa densité et sa fonction de distribution s'écrivent

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

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

- $P(x_1 < X < x_2)$ est proportionnelle par rapport la longueur de l'intervalle dans (a,b).

$$F(x_2) - F(x_1) = \frac{x_2 - x_1}{b-a} \quad si\ x_1, x_2 \in (a,b)$$

Loi uniforme

- Espérance

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

- Variance

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

- $U(0,1)$: nombre aléatoire => un outil important pour générer les variables aléatoires des autres lois de distribution

Loi normale

- Une variable aléatoire X suit une loi normale $N(\mu, \sigma^2)$ si sa densité s'écrit
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$
- Espérance
$$E(X) = \mu$$
- Variance
$$Var(X) = \sigma^2$$

Loi exponentielle

- Une variable aléatoire X suit une loi exponentielle de paramètre $\lambda > 0$ si sa densité et sa fonction de distribution s'écrivent
$$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0$$
$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$$
- Espérance
$$E(X) = \frac{1}{\lambda}$$
- Variance
$$Var(X) = \frac{1}{\lambda^2}$$
- Temps d'arrivée complètement aléatoire ou Temps de service très variable

Loi normale

- $\lim_{x \to -\infty} f(x) = 0$, $\lim_{x \to +\infty} f(x) = 0$
- Symétrique en μ: $f(\mu - x) = f(\mu + x)$
- Médiane = moyenne = μ
- Si $Z \sim N(\mu, \sigma^2)$ et $X \sim N(0,1)$ alors
$$X = \frac{Z - \mu}{\sigma}$$
$$F(x) = \Phi(x) = \int_{-\infty}^{x} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$$

Loi exponentielle

- Absence de mémoire

Proposition Une variable aléatoire X à valeurs dans $\mathbb{R}^+$, de fonction de répartition continue suit une loi exponentielle si et seulement si pour tout $t, h \geq 0$,
$$P(X + t + h | X > t) = P(X > h)$$

Proposition Une variable aléatoire X à valeurs dans $\mathbb{R}^+$ suit la loi $Exp(\lambda)$ si et seulement si pour tout $t >$ 0:
$$P(X \leq t + h | X > t) = \lambda h + o(h)$$

Loi exponentielle

- Stabilité

Considérons n variables aléatoires indépendantes, $X_1, \dots, X_n$ de lois respectives $Exp(\lambda_1), \dots, Exp(\lambda_n)$. Alors

$$Y = \min\{X_1, \dots, X_n\}$$

suit une lois exponentielles $Exp(\lambda_1 + \cdots + \lambda_n)$ et pour tout $i = 1, \dots, n$:

$$P(Y = X_i) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}$$

Loi lognormale

- Densité

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log(x)-\mu)^2}{2\,\sigma^2}}, \qquad x \geq 0$$

- Espérance

$$E(X) = e^{\mu+\sigma^2/2}$$

- Variance

$$Var(X) = e^{2\mu+\frac{\sigma^2}{2}}\left(e^{\sigma^2} - 1\right)$$

- Si $Y \sim N(\mu, \sigma^2)$ alors $X = e^Y \sim logN(\mu, \sigma^2)$

Loi de Weibull

- Densité

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x-\nu}{\alpha}\right)^{\beta-1} e^{\left(\frac{x-\nu}{\alpha}\right)^{\beta}}, \qquad x \geq \nu$$

- Trois paramètres
  - Location $-\infty < \nu < +\infty$
  - Largueur $\alpha > 0$
  - Forme $\beta > 0$
- Si X décrit la durée de vie d'un équipement
  - pour β < 1, le taux de défaillance est décroissant avec le temps (rodage, pannes de jeunesse)
  - pour β = 1, le taux de défaillance est constant et indépendant du temps (défauts aléatoires, loi exponentielle)
  - pour β > 1, le taux de défaillance est croissant avec le temps (phénomène d'usure, par exemple pour des roulements, moteurs…)
- Si $\beta = 1$, $X \sim Exp(\frac{1}{\alpha})$

Distribution empirique

- Échantillon : $X_1, X_2, \dots, X_n$ indépendants et identiquement distribués
- Cas impossible (ou non nécessaire) d'établir une variable aléatoire d'une loi paramétrique spécifique
- Avantage: sans hypothèse sauf les valeurs observées
- Désavantage: les observations ne couvent pas le support entier des valeurs possibles

Processus de comptage

- Processus de comptage

Soit $(N_t)_{t\geq 0}$ un processus stochastique à temps continu. On dit que N_t est un processus de comptage si pour la trajectoire de N_t est croissante par sauts d'amplitude 1.

Exemple Le nombre de clients arrivé dans l'intervalle de temps $[0, t]$.

Processus de sauts

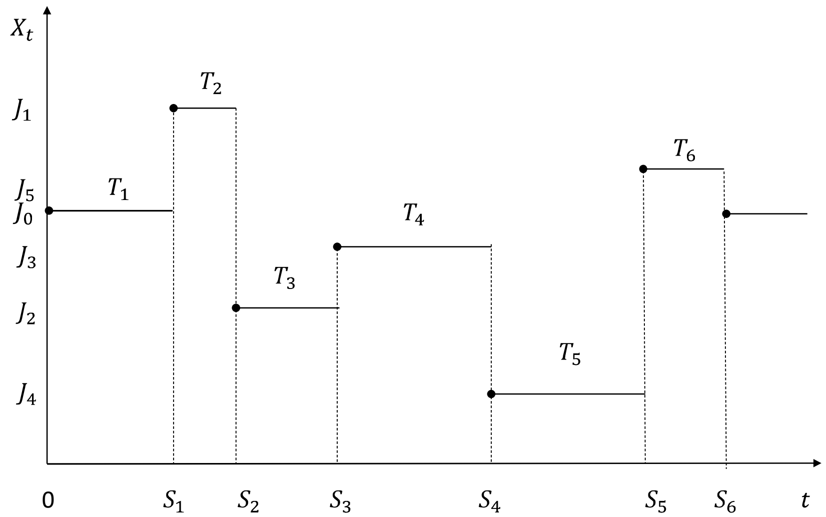
En supposant $S_0 = 0$, on a

- $X_t = J_{N_t}$
- $S_n = \sum_{i=1}^{n} T_i$
- $N_t = \sum_{n\geq 1} 1_{\{S_n \leq t\}}$
- Pour tout t ≥ 0, $N_t = n \iff S_n \leq t < S_{n+1}$

Processus de sauts

- Soit $(X_t)_{t\geq 0}$ un processus à temps continu à valeur s dans E, qui décrit l'état d'un système à l'instant t.
- La chaîne $(S_n)_{n\in\mathbb{N}}$ à valeurs dans $\mathbb{N}$, décrit les instants de sauts successifs de X_t avec
$$0 = S_0 < S_1 < \cdots < S_n < S_{n+1} < \cdots$$
- Le processus $(T_n)_{n\in\mathbb{N}}$ est la durée de temps entre chaque saut
$$T_n = S_n - S_{n-1}$$
- Soit $(J_n)_{n\in\mathbb{N}}$ la chaîne d'espace d'état E et J_n est l'état du système au n-ième saut au temps S_n.
- Soit $(N_t)_{t\geq 0}$ un processus décrit le nombre de sauts de X_t dans l'intervalle de temps $[0, t]$=>processus de comptage

Processus de sauts



Processus de Poisson

- Un processus de comptage N_t est appelé un processus de Poisson d'intensité λ si $(T_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires indépendantes et de même loi, exponentielle $Exp(\lambda)$.

Simulation de Monte Carlo

Soient $Z_1, Z_2, \dots, Z_n$ sont des variables aléatoires indépendantes et identiquement distribuées (i.i.d.) avec espérance $z = E(Z_1)$ et variance finie. La probabilité que la moyenne des échantillons soit proche de z est grande.

Quelque soit $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{Z_1 + Z_2 + \cdots + Z_n}{n} - z\right| < \varepsilon\right) = 1$$

Estimateur de

$$E\ (Z) = \overline{Z_n} = \frac{Z_1 + Z_2 + \cdots + Z_n}{n}$$

Estimateur de $P(Z_i > t)$

$$E(1_{\{Z_i>t\}}) = \frac{1_{\{Z_1>t\}} + 1_{\{Z_2>t\}} + \cdots + 1_{\{Z_n>t\}}}{n}$$

Processus de Poisson

- Pour $\mathrm{k} \in \mathbb{N}, 0 \le s \le t, \lambda > 0$, on a

$$P(N_t - N_s = k) = e^{\lambda(t-s)} \frac{\left(\lambda(t-s)\right)^k}{k!}$$

Alors $N_t - N_s$ suit une distribution de Poisson de paramètre $\lambda(t-s)$.

$$N_t - N_s \sim Pois(\lambda(t-s))$$

En particulier,

$$E(N_t - N_s) = \lambda(t-s)$$

Version infinitésimale:

$$P(N_{t+h} - N_t \ge 2) = o(h)$$
$$P(N_{t+h} - N_t = 1) = \lambda h + o(h)$$

Pour tous $t \ge 0$, N_t suit la loi de Poisson $Pois(\lambda t)$.

Simulation en Python

- Probabilité et statistiques
  - Module statistics
    - mean, median, mode, stdev, variance
  - Module random
    - random(), uniform(), randrage(),expovariate(),choice()
  - Module scipy.stats
    - rvs: générateur
    - pdf: densité
    - cdf : fonction de distribution
    - sff;: fonction de survie : 1-cdf
    - Ppf: fonction de fractile
    - Isf: fonction inverse de sff
    - stats: espérance, variance ,....
    - moment: moment non-central de la distribution