

Chapitre 2. Techniques d'évaluation des performances

Parmi les techniques principales d'évaluation des performances des systèmes (informatiques par exemple), on distingue :

1. Processus aléatoires (ou stochastiques) en particulier les chaînes de Markov.
2. Estimation statistique et tests d'hypothèses.
3. Optimisation, programmation mathématique et recherche opérationnelle.
4. Techniques graphiques.
5. Techniques de simulation.

Dans ce chapitre, nous détaillerons les deux premières techniques et renvoyons l'étudiant aux cours de recherche opérationnelle pour réviser les techniques du point 3.

1. Processus aléatoires.

Ce sont des outils de modélisation de la dynamique de phénomènes aléatoires, par exemple la dynamique d'un système dont l'évolution dans le temps est aléatoire.

On rappelle que dans les cours d'initiation au calcul de probabilités, on définit une variable aléatoire comme une variable qui peut prendre telle ou telle valeur à l'issue d'une expérience aléatoire.

Exemples :

1. lors du lancement d'une pièce de monnaie le résultat est imprévisible et on sait seulement que c'est « 0 » : Pile ou « 1 » : Face.
2. Le lancement d'un dé fournira une valeur de 1 à 6.

Par analogie, une **fonction aléatoire** est une fonction qui peut prendre telle ou telle allure à l'issue d'une expérience. L'aspect concret pris par la fonction à l'issue de l'expérience s'appelle **réalisation** ou **trajectoire**.

De manière plus rigoureuse, une variable aléatoire sera définie comme une application $X : W \rightarrow E$. L'ensemble $W = \{w\}$ représente l'ensemble des événements élémentaires (ou des issues possibles de l'expérience, Pile ou face dans l'exemple 1). (E est l'ensemble des valeurs possibles, par exemple 0 ou 1 dans l'exemple 1 ; ce peut être l'ensemble des nombres réels ou un de ses sous-ensembles dans le cas général).

Par analogie, une fonction aléatoire peut être définie comme une famille de variables aléatoires dépendant d'un paramètre t , $\{X(t), t \in T\}$. La fonction $X(t) = X(t, w)$ dépend en fait de deux paramètres : $w \in W$ représentant le hasard et $t \in T$, un paramètre ayant une interprétation donnée. L'ensemble des valeurs de X s'appelle espace des états ou des phases.

† si $T = \mathbb{R}$ ou \mathbb{R}^+ , t peut être interprété comme le temps est $X(t)$ est appelé **processus aléatoire** (ou **stochastique**) car on peut l'interpréter comme la valeur de l'état d'un système à

l'instant t . le but de l'analyse consiste à prédire la valeur de $X(t)$ (au sens de sa loi de probabilité) sachant l'observation du processus avant l'instant t .

Exemples de processus aléatoires :

- Signal (ou information) transmise par un canal de communication à l'instant t .
- Etat d'un processus (au sens informatique : séquence d'instructions) à l'instant t : actif, bloqué ou en état de réflexion.
- Nombre de messages dans une mémoire à l'instant t .
- Nombre de processus (ou jobs) en cours d'exécution par le processeur à t .

L'ensemble E s'appelle **espace des états** (ou des **phases** en relation avec les modèles de signaux en électricité).

† Si $t \in \mathbb{N}$ (le temps est discret : $t=0,1,2,\dots$), la fonction aléatoire s'appelle **suite (ou chaîne) aléatoire** (c'est le cas pour le temps naturel, heure, jour, mois, ...ou lorsqu'on discrétise le temps réel pour l'implémenter sur ordinateur en slots).

Exemples :

- volume de la production d'une usine le mois t .

† Si $t \in \mathbb{R}^3$ (ou \mathbb{R}^n) on parle de champ aléatoire.

Exemples:

- Position d'une molécule dans un gaz
- Représentation d'une image en 2D ou 3D (vision ou reconnaissance de formes).
- Epicentre d'un séisme qui est caractérisée non seulement par sa position en 3D (latitude, longitude, profondeur, mais également son amplitude : $t \in \mathbb{R}^4$.

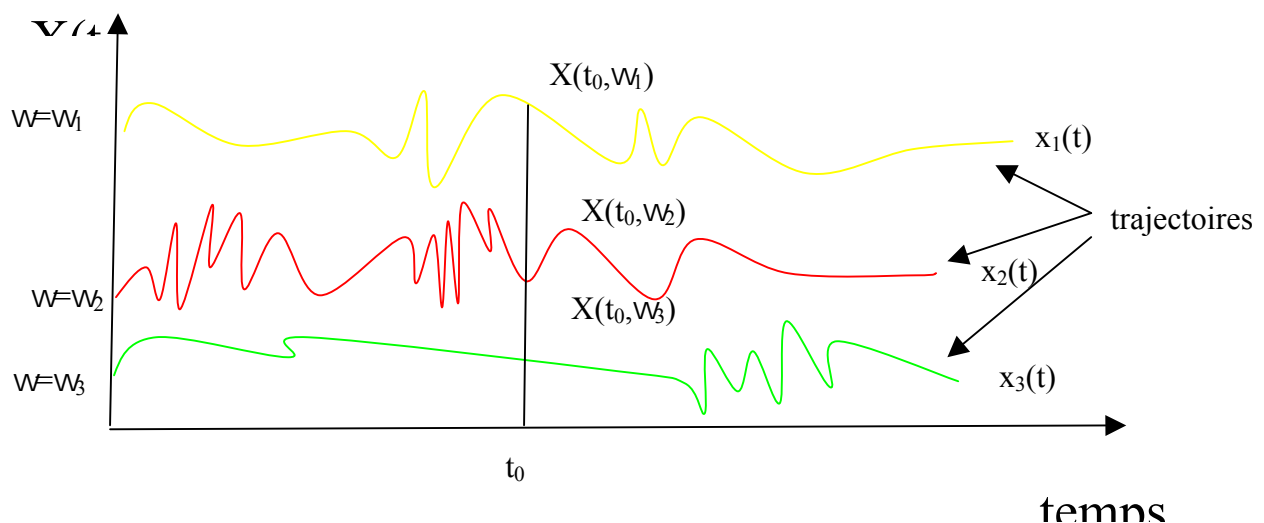


Figure 1. Réalisations (ou trajectoires) du

On s'intéressera principalement aux processus aléatoires, les méthodes décrites ci-dessous sont similaires dans le cas de fonctions aléatoires.

Comment définir un processus aléatoire ?

On rappelle que pour définir une variable aléatoire il suffit de connaître sa fonction de répartition (ou sa fonction de densité si elle existe).

Pour un processus aléatoire, la donnée de la fonction de répartition à l'instant t , $F_t(x) = P\{X(t) \leq x\}$ ou de la fonction de densité $f_t(x) = F'_t(x)$ est insuffisante. Il faut pour cela donner toutes les fonctions de répartition fini-dimensionnelles (f.r.f.)

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = P\{X(t_1) \leq x_1, \dots, X(t_n) \leq x_n\} \quad (1)$$

" $t_1 < t_2 < \dots < t_n$, " x_1, \dots, x_n , " $n=1, 2, \dots$ Ces f.r.f. doivent vérifier les conditions de symétrie

$$F_{t_{i_1}, t_{i_2}, \dots, t_{i_n}}(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)$$

pour tous $i_1, \dots, i_n \in \{1, 2, \dots, n\}$, et les conditions de compatibilité

$$F_{t_1, \dots, t_n, t_{n+1}, \dots, t_m}(x_1, x_2, \dots, x_n, \infty, \dots, \infty) = F_{t_1, \dots, t_n}(x_1, \dots, x_n)$$

Caractéristiques probabilistes des processus aléatoires.

Il est difficile en général de connaître les loi fini-dimensionnelles (trop complexes), à l'exception de cas particuliers (cas gaussien). On peut cependant obtenir une information partielle fournie par les moments (moyenne, dispersion, corrélation) et qui sont suffisants pour les études pratiques.

Moments.

$$m_{j_1, j_2, \dots, j_p}(t_1, t_2, \dots, t_p) = E\{X(t_1)^{j_1} X(t_2)^{j_2} \dots X(t_p)^{j_p}\}$$

Fonction moyenne (ou espérance mathématique).

C'est une fonction déterministe (non aléatoire) mais dépendante du temps t

$$m(t) = E\{X(t)\} = \int x f_t(x) dx$$

Fonction variance $\sigma^2(t) = \text{Var}\{X(t)\} = E\{[X(t) - m(t)]^2\}$

Fonction de covariance : C'est la covariance entre les valeurs $X(t)$ et $X(s)$,

$$K(t,s) = \text{Cov}[X(t), X(s)] = E\{[X(t) - m(t)][X(s) - m(s)]\} = E\{X(t)X(s)\} - E\{X(t)\}E\{X(s)\}$$

Fonction de corrélation (ou d'autocorrélation)

$$\rho(t,s) = \frac{K(t,s)}{\sqrt{\text{Var}[X(t)] \cdot \text{Var}[X(s)]}}$$

Propriétés.

- $\rho(t,s)$ mesure le degré de dépendance linéaire entre les valeurs $X(t)$ et $X(s)$ du processus.
- Si $\rho(t,s) = 0$ alors on dit que $X(t)$ et $X(s)$ sont non corrélées.
- $K(t,t) = \text{Var}[X(t)] = \sigma^2(t)$ = variance de la variable $X(t)$ et $\rho(t,t) = 1$
- $K(t,s) = K(s,t)$ (symétrie).
- $|K(t,s)| \leq \sqrt{\text{Var}[X(t)]} \sqrt{\text{Var}[X(s)]}$ (inégalité de Cauchy-Schwartz) et $0 \leq \rho(t,s) \leq 1$.
- La fonction $K(t,s)$ est définie positive.

Classes particulières de processus aléatoires.

Dans certains cas, la définition (1) du processus aléatoire peut être simplifiée.

- (i) si $t_1 < t_2 < \dots < t_n < t_{n+1} < \dots$, les variables aléatoires (accroissements) **Processus à accroissements indépendants**. On dit qu'un processus est à accroissement indépendants

$$U_1 = X(t_2) - X(t_1), U_2 = X(t_3) - X(t_2), U_3 = X(t_4) - X(t_3), \dots, U_k = X(t_{k+1}) - X(t_k),$$

sont indépendantes.

Exemple 1: Le processus de **Poisson** à valeurs dans $\mathbb{N} = \{0, 1, 2, \dots\}$ de paramètre λ est un processus à accroissements indépendants tel que

- $P(X(0)=0)=1$.
- les accroissements $X(t)-X(s)$, $0 \leq s \leq t$ suivent une loi de Poisson de paramètre $\lambda(t-s)$:

$$P\{X(t)-X(s)=k\} = \frac{[\lambda(t-s)]^k}{k!} e^{-\lambda(t-s)}, k=0, 1, 2, \dots$$

Exercice 1: (a) Vérifier que pour le processus de Poisson : $m(t) = E\{X(t)\} = \lambda t$; $\text{Var}\{X(t)\} = \lambda t$.
(b) Calculer $K(t,s)$.

Le processus de Poisson se rencontre dans de nombreuses applications. Il peut modéliser :

- le nombre de pannes d'un système informatique durant $(0,t)$.
- le nombre de visites d'un site Web durant $(0,t)$.
- le nombre de particules enregistrées par un compteur Geiger....

(iv) le nombre d'appels d'abonnés à un central téléphonique durant $(0,t)$.

© Notons qu'on peut définir le processus de Poisson d'une autre manière. Au lieu de compter le nombre d'événements (pannes, visites, appels...) durant la période $(0,t)$, on peut d'une manière alternative s'intéresser aux instants même d'occurrence de ces événements : $t_1 < t_2, \dots < t_n < t_{n+1}$ où t_n est l'instant du n -ème événement.

Si $\{x_n = t_{n+1} - t_n; n=1,2,\dots\}$ sont des variables aléatoires indépendantes de même loi $F(x) = P\{x_n < x\}$ (indépendante de n), alors on parle de **processus de renouvellement**. L'interprétation provient de l'exemple (i) ci-dessus: on suppose que le système est « renouvelé » ou « réparé » immédiatement après la panne ; si on suppose que la durée de renouvellement est négligeable (devant les intervalles entre pannes), alors t_n représente aussi l'instant du n -ième renouvellement. Un processus de renouvellement tel que la loi des intervalles entre pannes est exponentielle $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$ n'est rien d'autre qu'un processus de Poisson de paramètre λ .

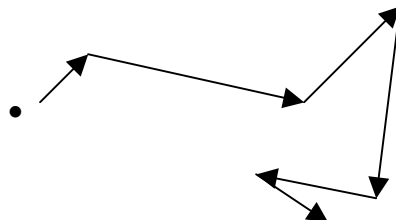
On a donc deux manières de définir le processus de Poisson : soit en comptant le nombre d'événements durant $(0,t)$; soit en considérant les instants même d'occurrence de ces événements. On peut en déduire deux interprétations possibles du paramètre λ :

-la première définition, montre que $\lambda = \{X(1)\}$ i.e. λ peut-être interprété comme le nombre moyen d'événements (pannes, appels...) durant la période $(0,t)$.

-la seconde définition par contre indique que $E\{x_n\} = 1/\lambda$ i.e. λ est aussi l'inverse du temps qui sépare les occurrences de deux événements successifs.

Exemple 2 : Un processus **Gaussien** est un processus à accroissements indépendants tel que les accroissements $X(t) - X(s)$ suivent une loi normale.

En particulier, le processus de **Wiener** est un processus Gaussien tel que les accroissements $X(t) - X(s)$ suivent une loi normale $N(0; s(t-s))$ $s \leq t$ de moyenne nulle. Ce processus a été utilisé notamment pour modéliser le **mouvement Brownien** : le mouvement chaotique de particules (molécules) dans un corps liquide suite à un grand nombre de « petits » chocs. C'est également un « bon » modèle du **processus d'erreur** qu'on appelle également **bruit blanc** (en relation avec la théorie du signal en électronique).



(ii). **Processus stationnaire**. Le processus est **stationnaire au sens strict** (ou au sens faible) si les f.r.f. sont invariantes par translation dans le temps i.e.

$(X(t_1), X(t_2), \dots, X(t_n))$ et $(X(t_1+h), X(t_2+h), \dots, X(t_n+h))$ ont même distribution " t_1, \dots, t_n , " $h > 0$, " n .

En particulier, si $X(t)$ est strictement stationnaire, alors on a

1. La moyenne $m(t)=E\{X(t)\}$ ^a $m=\text{constante}$ indépendante de t .

2. La fonction de corrélation $K(t,s)=f(t-s)$ est une fonction qui ne dépend ni de t , ni de s , mais seulement de la différence $t-s$.

La stationnarité stricte est en général difficile à vérifier, aussi pour la majorité des applications on se contente d'un autre type de stationnarité. On dit qu'un processus est **stationnaire au sens large ou à l'ordre 2** si les conditions 1 et 2 sont vérifiées.

Certains processus peuvent être définis uniquement à partir de leurs accroissements. On dira alors qu'un processus est à **accroissements stationnaires** si les variables (les accroissements) $X(t_1)-X(t_0), X(t_2)-X(t_1), \dots, X(t_n)-X(t_{n-1})$ sont invariants (en loi) par translation dans le temps.

(ii) **Série chronologique (ou temporelle)**: Ce processus est représenté sous la forme :

$$X(t)=X_{t+1}+X_{t+2}+\dots+X_{t+k}+\epsilon(t)$$

où $\epsilon(t)$ est un bruit blanc. Ce type de processus se rencontre fréquemment en théorie du signal. L'appellation provient surtout d'application économique où le temps t est mesuré en années: -indices des prix ; -nombre d'étudiants accédant à l'université. -évolution des salaires ; -consommation des ménages....

(iii) **Processus markovien**. On dit qu'un processus est markovien ou sans mémoire, si son évolution future ne dépend pas du passé, mais seulement du présent (propriété de Markov) i.e. " $t_1 < t_2 < \dots < t_n < t_{n+1} < \dots$, " n .

$$P\{X(t_{n+1}) \leq x_{n+1} / X(t_0)=x_0, \dots, X(t_n)=x_n\} = P\{X(t_{n+1}) \leq x_{n+1} / X(t_n)=x_n\}$$

Exemple : Les processus de Poisson, Gaussiens, à accroissements indépendants sont des processus markoviens.

Si le temps est discret, on parle de chaîne de Markov. L'étude de tels processus est plus simple car ne nécessitent pas la connaissance de toutes les distributions fini-dimensionnelles. Les domaines d'application sont ainsi très variés, en particulier pour l'étude des systèmes informatiques ; algorithmes stochastiques dans les problèmes d'Intelligence artificielle, problèmes de reconnaissance de formes et de vision (réseaux de neurones, etc...

3. Techniques statistiques.

Ce sont des techniques d'inférence permettant d'estimer les paramètres ou de déterminer expérimentalement les variables ou processus aléatoires à partir d'observations empiriques (dites échantillons).

Exemples : Si on sait que le nombre de pannes d'un système suit un processus de Poisson, ces techniques permettent d'estimer le paramètre λ . Si on n'a aucune information sur la loi, ces techniques permettent également de trouver parmi les lois possibles celle qui ajuste au mieux les données des observations du processus durant une certaine période.

4. Méthodes d'optimisation.

Lors de l'étude d'un système il ne s'agit pas seulement d'évaluer les performances du système, mais également de les optimiser : par exemple minimiser le temps de réponse, maximiser le débit ou la disponibilité etc... Il s'agit alors d'optimiser une certaine fonction dépendant de plusieurs paramètres $f(x_1, x_2, \dots, x_n)$ avec éventuellement un certains nombres de contraintes sur les valeurs x_1, \dots, x_n . Les méthodes de programmation mathématique s'intéresse à divers classes de fonctions f et sont suffisamment détaillée dans les cours de recherche opérationnelle.

5. Techniques de simulation.

Comme cela a été évoqué, la simulation consiste à générer la dynamique du système (décrite par un certain processus aléatoire), artificiellement sur ordinateur (à l'aide d'un programme) plutôt qu'avec une méthode analytique. Cette dynamique du processus aléatoire est souvent décrite à l'aide de certains événements et des durées. On génère ainsi plusieurs réalisations (trajectoires) et on effectue un traitement statistique des mesures de performance qui nous intéresse. Ce type de simulation est souvent appelée simulation à événements discrets, car la réalisation du processus consiste en une succession d'événements à générer. Les techniques graphiques permettent de doter le programme de simulation d'un module d'animation qui permet de réaliser rapidement différents scénarios et de les comparer. Ces techniques seront détaillées dans les chapitres suivants.

Travaux dirigés

Série N°1

Exercice N°1. Soit le processus $X(t) = 1 + Ut + Vt^2$, où U et V sont deux variables aléatoires non corrélées ($\text{Cov}(U, V) = 0$) ; U suit une loi normale $N(0, 3)$ et V une loi normale $N(0, 1)$. Calculer les fonctions moyenne, d'auto covariance et d'auto corrélation. En déduire si le processus $\{X(t)\}$ est stationnaire au sens large ? au sens strict ?

Exercice N°2. Soient $\{X_1(t)\}$ et $\{X_2(t)\}$ deux processus de Poisson de taux $\lambda > 0$. On définit le processus $X(t)$, $t \in \mathbb{R}$ comme suit :

$$X(t) = -X_1(-t) \quad \text{si } t < 0$$

$$X(t) = X_2(t) \quad \text{si } t \geq 0$$

1. Calculer les fonctions moyennes $E\{X(t)\}$ et de covariance $K(s, t) \forall s, t \in \mathbb{R}$.

2. En déduire si le processus $\{X(t)\}$ est stationnaire au sens large ? au sens strict ?

Exercice 3 : Soient deux variables aléatoires U et V de densités $f_U(x) = \lambda e^{-\lambda x}$, et $f_V(x) = \mu e^{-\mu x}$ ($x \geq 0$), indépendantes.

1. Ecrire la densité unidimensionnelle du processus aléatoire $X(t) = U + Vt$ pour $t > 0$.
2. Calculer les fonctions moyenne, variance, d'autocovariance et d'autocorrélation de ce processus.

3. En déduire si le processus est stationnaire : au sens large ? au sens strict ?