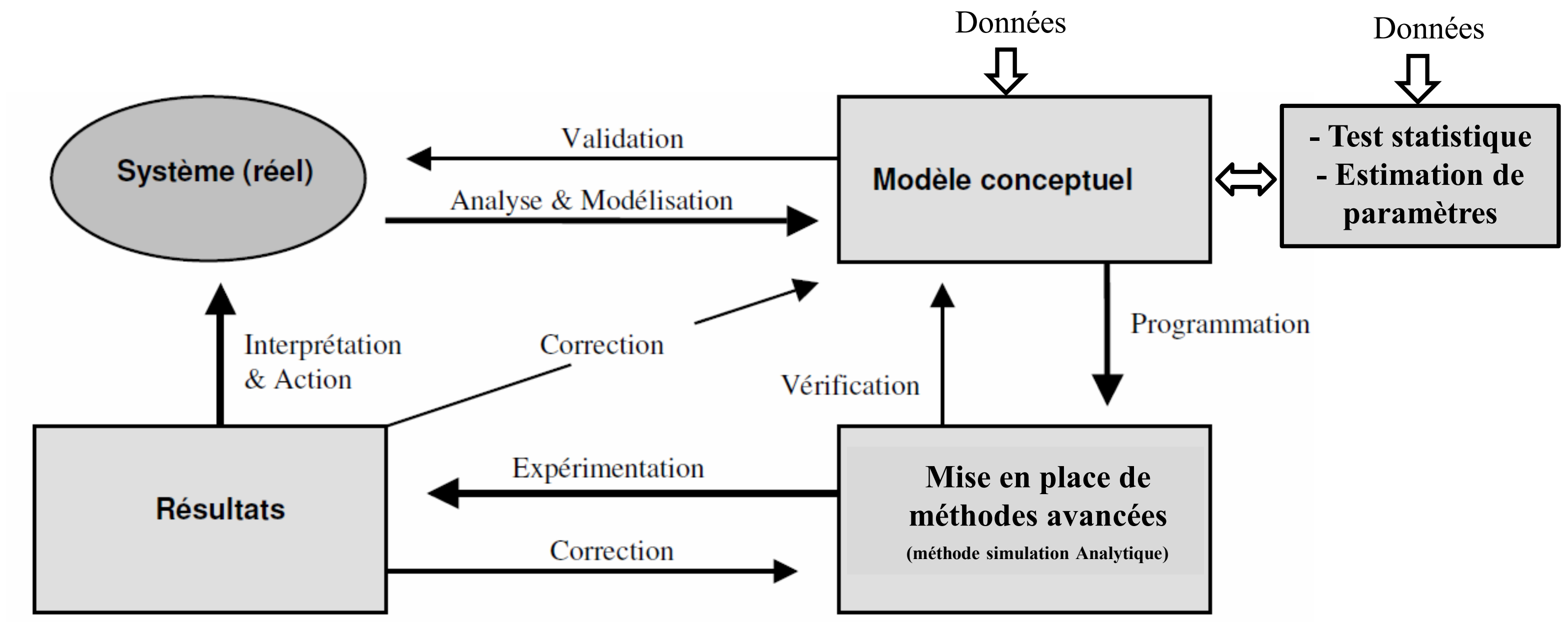
Evaluation de performances

Fondements mathématiques pour l'évaluation de performances

- Variables aléatoires
- Quelques lois de probabilité
- Intervalle de confiance
- Tests statistiques
- Notions de processus stochastiques

Phuc Do
phuc.do@univ-lorraine.fr

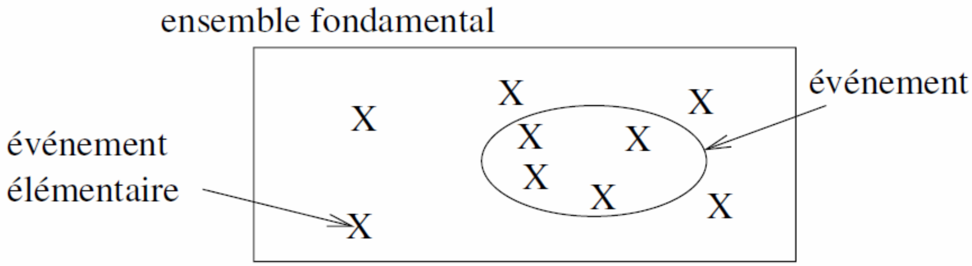
Vision générale sur l'évaluation de performances par des méthodes avancées



Besoins:

- Modélisation
- Estimation des paramètres
- Simulation de comportements
- Calcul des indicateurs de perf.
- Interprétation de résultats obtenus
- …

- Expérience aléatoire, événement
  - Une ***expérience aléatoire*** est une expérience dont le résultat n'est pas prévisible
  - Le résultat d'une expérience aléatoire est appelée ***événement élémentaire***, ou encore épreuve ou réalisation.
  - La réunion de tous les événements élémentaires que peut produire une expérience aléatoire est appelée ***ensemble fondamental***, ou encore espace des épreuves (notation : Ω).
  - Tout sous-ensemble de l'ensemble fondamental est un ***événement***



événements et ensemble fondamental

- Concept de variables aléatoires
  - Une v.a est une variable qui peut prendre lors d'une expérience aléatoire une valeur quelconque, inconnue d'avance, parmi un ensemble de valeurs possibles, appelé espace d'état, noté E.
  - Une v.a X est parfaitement caractérisée par sa **fonction de répartition** F_X qui est l'application de R dans [0, 1] définie par $F_X(x) = P(X \leq x)$

- 2 types de variables aléatoires
  - v.a discrète est une v.a X telle que l'ensemble des valeurs x_i qu'elle peut prendre est **dénombrable**
  - v.a continue est une v.a X telle que l'ensemble des valeurs qu'elle peut prendre n'est **pas dénombrable**

- V.a **discrètes**
  - La loi de probabilité d'une variable aléatoire discrète *X* est la fonction définie par

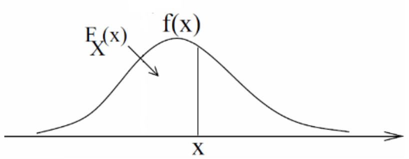
$$x \longmapsto p_i = P[X = x_i] \text{ où } \sum_{i \in \mathbb{N}} p_i = 1$$

  - La fonction de répartition: $$F(x) = \sum_{x_i \leq x} p(x_i)$$

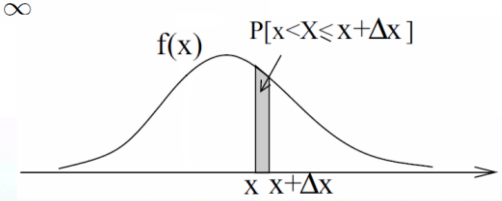
Ex: *Etats d'un système: marche, panne, dégradé, …*

- V.a **continues**
  - Une v.a absolument continue est une v.a X dont la fonction de répartition $F_X(x)$ est absolument continue. Il existe alors une fonction f_X, appelée densité de probabilité de la variable aléatoire X, telle que:

$$\forall x \in \mathbb{R} \quad F_X(x) = \int_{-\infty}^{x} f_X(t)d$$



  - La densité de probabilité: $$f(x) = \frac{dF(x)}{dx}$$ avec $$\int_{-\infty}^{+\infty} f(x)dx = 1$$



Ex: *Date de panne, niveau de dégradation,…*

Moments d'une v.a

- Espérance mathématique:
  - L'espérance mathématique d'une v.a X est, si elle existe, la moyenne des valeurs possibles de X pondérées par les probabilités correspondantes
    - *Si la variable est discrète:*

$$E\{X\} = \sum_{i\in I} x_i p_i$$

    - *Si la variable est continue:*

$$E\{X\} = \int_{\mathbb{R}} x f(x) dx$$

- Variance:
  - Une mesure servant à caractériser la dispersion d'une distribution ou d'un échantillon
  - La variance d'une variable aléatoire X est: $\sigma^2_X = E[X-E(X)]^2 = E(X^2)-[E(X)]^2$
    - *Si la variable est discrète:*

$$\sigma_X^2 = \sum_{i\in I} (x_i - E\{X\})^2 p_i$$

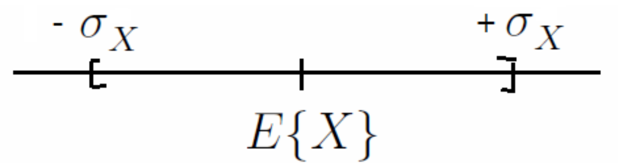
    - *Si la variable est continue:*

$$\sigma_X^2 = \int_{\mathbb{R}} (x - E\{X\})^2 f(x) dx$$

  - L'écart-type, ou déviation standard: σ_X

Moments d'une v.a

- Variance:
  - ***La variance donne une indication sur la dispersion, ou "étalement", de la distribution autour de la valeur moyenne***



- Coefficient de variation
  - Le coefficient de variation mesure la dispersion relative de la variable X par rapport à sa moyenne

$$\mathrm{cv}_X = \frac{\sigma_X}{E[X]}$$

Quelques lois de probabilité

- **Objectif:**
  - Analyser des données (mesures) disponibles
  - Simuler des données
  - Base de processus stochastiques

Variables aléatoires discrètes

- Loi uniforme discrète
  - C'est la loi de probabilité discrète à valeurs dans {x1, . . . , xn} telle que

$$P[X=k]=1/n \quad \forall k \in [1,n]$$

  - On a : $E\{X\}=\dfrac{n+1}{2}$ $\qquad \sigma_X^2 = \dfrac{n^2-1}{12}$

Sous Matlab: randi([a,b])

- Loi de Bernoulli
  - C'est la loi de probabilité discrète notée B(p) à valeurs dans {0, 1} telle que

$$P(X=1)=p \quad \text{et} \quad P(X=0)=1-p$$

$$E\{X\}=p \qquad \sigma_X^2 = p(1-p)$$

  - Ex: *p*=0,5: X=1 = «Pile»; X=0 = «Face»

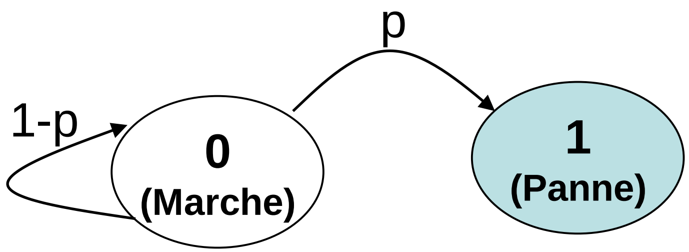
Variables aléatoires discrètes

- Loi géométrique: la probabilité, lors d'une succession d'épreuves de Bernoulli indépendantes, d'obtenir *k* échecs suivi d'un succès

$$P(X = k) = (1 - p)^k p$$

- Avec $E\{X\} = \frac{1}{p}$ $\sigma_X^2 = \frac{1-p}{p^2}$

- Ex: la probabilité qu'un component tombe en panne dans une journée est *p=0,2*.



1- Quelle est la probabilité que le composant tombe en panne à la *n-ème* journée?

2- Déterminer la durée de vie moyenne du composant ?

Variables aléatoires discrètes

Loi de Poisson

- Une variable aléatoire X discrète suit une loi de Poisson de paramètre λ, et est notée P(λ), si l'ensemble des valeurs k qu'elle peut prendre est N et

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \quad \forall k\in\mathbb{N}, \quad \text{et où } \lambda\in\mathbb{R}^{+*}$$

- Fonction de réparation: $F(x)=\sum_{k=0}^{x}\frac{\lambda^k}{k!}e^{-\lambda}$

- On a: $E\{X\}=\lambda$ $\qquad \sigma_X^2=\lambda$

Exemple:

Le nombre moyen de requêtes envoyées à un serveur est de 10 requêtes/seconde. Sachant que le nombre de requêtes envoyées en une seconde suit une loi de Poisson. Déterminer:

- la probabilité que 5 requêtes soient envoyées en une seconde ?

P[X=5] = 0,0378

sous Matlab: poisspdf(5,10)

- la probabilité que le nombre de requêtes envoyées en une seconde soit inférieur ou égale à 10

F(10)=0,5830

sous Matlab: poisscdf(10,10)

Variables aléatoires discrètes

- Loi empirique
  - une loi ou une formule issue de faits expérimentaux, ou validée par l'expérience
  - Fonction de masses: $P(X = x_i) = p_i$ pour $i = 1, \dots, m$
  - Fonction de répartition: $F(x_i) = \sum_{j=1}^{i} p_j$

Exemple:

Dans un atelier de production, on étudie X, le nombre de pannes par jour. Pour cela, on fait 100 expériences d'une journée. On obtient les résultats suivants :

Nombre de pannes x_i	0	1	2	3	4	5
fréquence n_i	21	38	22	10	5	4

- Déterminer la loi empirique de X ?

Variables aléatoires continues

- Loi uniforme contitue
  - Une variable aléatoire continue suit une loi uniforme sur un intervalle [*a, b*]

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si} \quad x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{si} \quad x < a \\ \frac{x-a}{b-a} & \text{si} \quad a \leq x \leq b \\ 1 & \text{si} \quad x > b \end{cases}$$

$$E\{X\} = \frac{1}{2}(a+b) \qquad\qquad \sigma_X^2 = \frac{(b-a)^2}{12}$$

Sous Matlab:

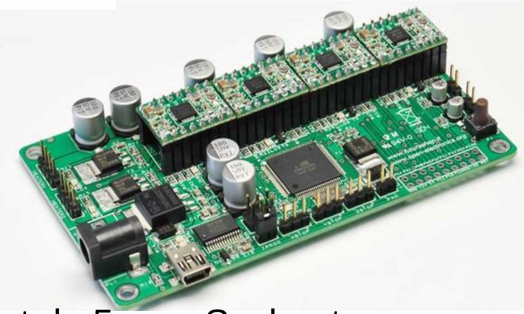
- rand() pour a =0 et b=1
- (b-a)*rand + a pour a et b quelconques

Variables aléatoires continues

- Loi exponentielle
  - Une variable aléatoire X continue suit une loi exponentielle de paramètre λ, et est notée E(λ), si

$$f(x) = \lambda e^{-\lambda x}\, U(x), \quad \text{où } \lambda \in \mathbb{R}^+ \quad \text{et } U(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$$

$$F(x) =$$



$$E\{X\} = 1/\lambda \qquad \sigma_X^2 = 1/\lambda^2$$

Exemple:

Le temps moyen de la première défaillance d'un composant électronique est de 5 ans. Sachant que son temps de tomber en panne suit une loi exponentielle. Déterminer:

- la probabilité que ce composant soit défailli avant 3 ans ?

F(3)=0,4512

Sous matlab: expcdf(3,5)=0,4512

Variables aléatoires continues

Loi Weibull

- Une variable aléatoire X continue suit une loi weibull de paramètres α et λ

$$f(t) = \begin{cases} \alpha\lambda^{\alpha}t^{\alpha-1}e^{-(\lambda t)^{\alpha}} & \text{pour } t > 0 \\ 0 \quad \text{sinon} & \end{cases}$$

α : paramètre de forme, λ : paramètre d'échelle

$$F(x) =$$

$$E\{X\} = \frac{1}{\lambda}\Gamma\left(\frac{1}{\alpha}+1\right) \qquad \sigma_X^2 = \frac{1}{\lambda^2}\left(\Gamma\left(\frac{2}{\alpha}+1\right) - \Gamma^2\left(\frac{1}{\alpha}+1\right)\right)$$

$$(\text{ fonction Gamma } \quad \Gamma(a) = \int_0 x^{a-1}\cdot e^{-x}dx \ )$$

- Si α=1 => loi exponentielle

Exemple:

Supposons que la défaillance d'un composant mécanique suit une loi Weibull de parametres (α =5, λ =2). Déterminer:

- la probabilité que ce composant soit défailli avant 3 ans ?

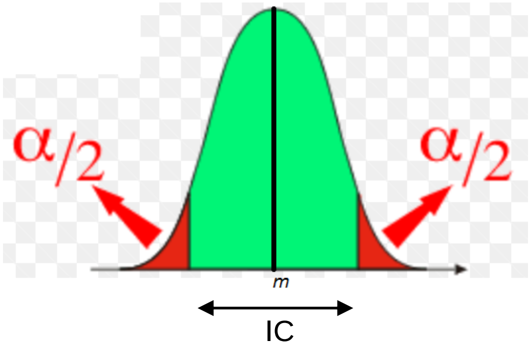
Sous Matlab: wblcdf(3,5,2)=0,302

Intervalle de confiance

Intervalle de confiance

- Soient : X une v.a de loi paramétrée par θ et $X_1,...,X_n$ n variables i.i.d selon la loi de X

- **Définition:** On appelle intervalle de confiance de niveau de confiance $1-\alpha$ du paramètre θ tout intervalle *IC* tel que : $P(IC \ni \theta) = 1-\alpha$ pour $\alpha \in [0,1]$ fixé.



❖ Remarques:

- Les bornes de l'intervalle de confiance *IC* dépendent de l'***échantillon***, elles sont donc aléatoires
- si α ***augmente*** (ou que si *n* augmente), l'amplitude de l'intervalle de confiance ***diminue***

i.i.d=indépendantes identiquement distribuées

Intervalle de confiance

❑ Données: 2 cas envisagés:

- a variable aléatoire mesurée est normale et le nombre de réalisations est quelconque
- la variable aléatoire mesurée n'est pas normale et le nombre de réalisations est important. Dans ce cas, la distribution de statistique (ex: la moyenne, variance, ..) tend vers une loi normale d'après le théorème central limite. On parlera d'intervalle de confiance asymptotique.

❑ Théorème limite centrale:

- Considérons un échantillonnage de taille N
- Plus N est grand, plus la distribution d'échantillonnage de la moyenne s'apparente à une distribution normale

$$\underbrace{\bar{X}}_{\text{Distribution d'échantillonnage de la moyenne}} \sim N\left(\mu_x, \underbrace{\frac{\sigma_x}{\sqrt{N}}}_{\text{erreur-type}}\right)$$

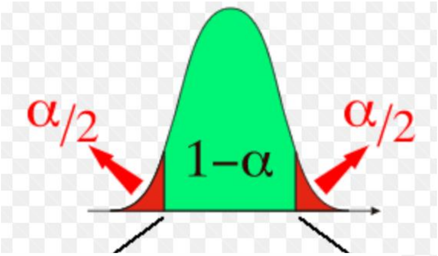
Intervalle de confiance pour l'espérance

Considérons $X \sim N(m,\sigma^2)$ et $X_1,\ldots,X_n$ n variables i.i.d selon la loi de X.

- **Cas où la variance est connue** $IC = \left[\widehat{m}_n - u\frac{\sigma}{\sqrt{n}}\ ,\ \widehat{m}_n + u\frac{\sigma}{\sqrt{n}}\right]$

où u est le fractile d'ordre $1-\frac{\alpha}{2}$ de la loi $N(0,1)$, i.e. $P(X \leq u) = 1-\alpha/2$

moyenne empirique: $\widehat{m}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$



- **Cas où la variance est inconnue** $IC = \left[\widehat{m}_n - t_\alpha\frac{\widehat{s}_n}{\sqrt{n}},\ \widehat{m}_n + t_\alpha\frac{\widehat{s}_n}{\sqrt{n}}\right]$

variance empirique modifiée: $\widehat{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \widehat{m}_n\right)^2$

où t_α est le fractile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student à *n-1* degrés de libertés

Ex: $n = 100 \Rightarrow t_\alpha = 1.98$

❖ ***Remarque:*** quand $n \to \infty$ (n≥30), on approxime la loi de Student par la loi normale centrée réduite. On retrouve alors le cas précédent

Intervalle de confiance pour la variance

- **Cas où l'espérance est connue**

$$IC(\sigma^2) = \left[n\frac{S_n^{*2}}{\chi^2_{1-\frac{\alpha}{2}}} \;,\; n\frac{S_n^{*2}}{\chi^2_{\frac{\alpha}{2}}} \right]$$

$$S_n^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$

où χ^2_α est le fractile d'ordre α de la loi $\chi^2(n)$

- **Cas où l'espérance est inconnue**

$$IC(\sigma^2) = \left[(n-1)\frac{\hat{S}_n}{\chi^2_{1-\frac{\alpha}{2}}} \;,\; (n-1)\frac{\hat{S}_n}{\chi^2_{\frac{\alpha}{2}}} \right]$$

où χ^2_α est le fractile d'ordre α de la loi $\chi^2(n-1)$

Tests statistiques

- Test du X2
- Test de Kolmogorov

Tests statistiques

- Un test d'hypothèse est une démarche consistant à rejeter ou à ne pas rejeter une hypothèse statistique, appelée *hypothèse nulle* (ou hypothèse zéro), noté H0, en fonction d'un jeu de données (échantillon),

- **Types de test:**
  - ***Tests paramétriques*** lorsque l'on stipule que les données sont issues d'une distribution paramétrée
  - ***Tests non paramétriques*** ne font aucune hypothèse sur la distribution sous-jacente des données
    - test du X^2
    - Test de Kolmogorov
    - …

Tests statistiques

- **Test du X^2 :**
  - H0 suppose que les données considérées proviennent de variables aléatoires discrètes suivant une loi de probabilité donnée
  - On souhaite tester la validité de cette hypothèse selon un niveau de confiance (par défaut 95%).
  - Sous Matlab: chi2test

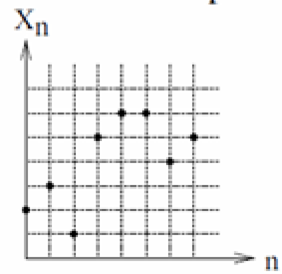
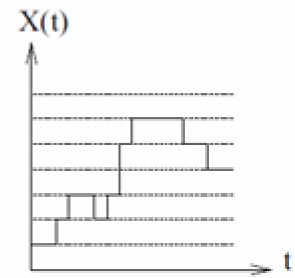
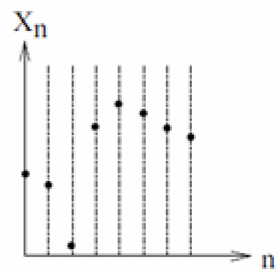
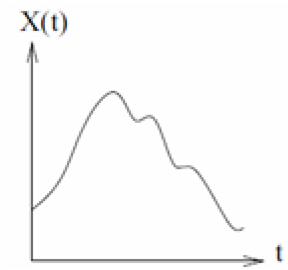
- **Test de Kolmogorov**
  - H0 suppose que les données considérées proviennent de variables aléatoires continues suivant une loi de probabilité donnée
  - On souhaite tester la validité de cette hypothèse selon un niveau de confiance
  - Sous Matlab: kstest ou kstest2

Notion de processus stochastique

Notion de processus stochastique

- Un processus stochastique $\{X(t)\}_{t \in T}$ est **une famille de variables aléatoires** X(t) où t est un paramètre
- Pour chaque valeur de t **l'état du processus** est défini par la valeur prise par la variable aléatoire X(t).
- **L'espace des paramètres** *T* est l'ensemble des valeurs que le paramètre peut prendre
- Lorsque T est **continu**, le processus est noté {X(t)} et lorsqu'il est **discret**, le processus est souvent noté $\{X_n\}$.
- **L'espace des états** ***E*** est l'ensemble des valeurs que les variables aléatoires peuvent prendre

Notion de processus stochastique

	T discret	T continu
E discret	**processus à temps discret et à espace d'état discret** ou chaîne à temps dicret  Ex: Nombre de requêtes moyen dans le système suivant chaque minute	**processus à temps continu et à espace d'état discret** ou chaîne à temps continu  Ex: nombre de clients dans le sytème à chaque instant *t*
E continu	**processus à temps discret et à espace d'état continu**  Ex: taux de retransimissions en fonction du jour	**processus à temps continu et à espace d'état continu**  Ex: processus de dégradation