MAINTENANCE PRÉDICTIVE

Plan du cours

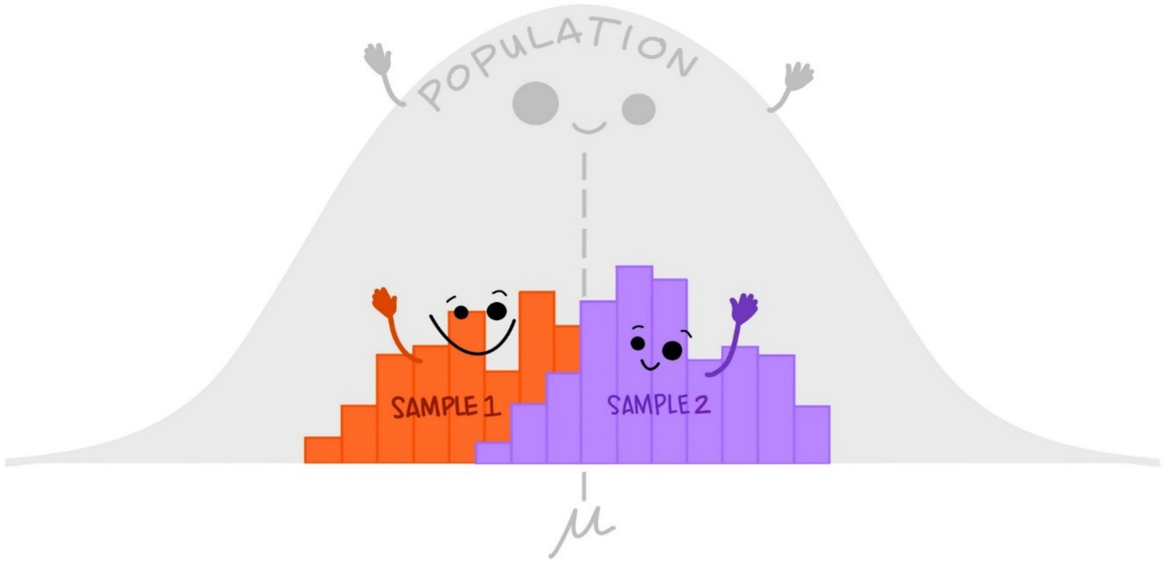
- Introduction
- Modélisation statistique des lois de vie
- Construction d’indicateurs de santé
- Sélection de variables, types de variables et prédiction
- Capteurs, analyse de risques, réseaux bayésiens
- Examen

STATISTIQUES

Modélisation statistique

- **Rappels:**
  - Les enjeux de la maîtrise de la maintenance :
    - Réduire les coûts, et ce, sous certaines contraintes
  - Vu en TP :
    - Les systèmes tombent en panne selon certaines probabilités
  - L’objectif de cette séance :
    - Quelles modélisations probabilistes adopter ?
    - Comment apprendre ces modèles ?

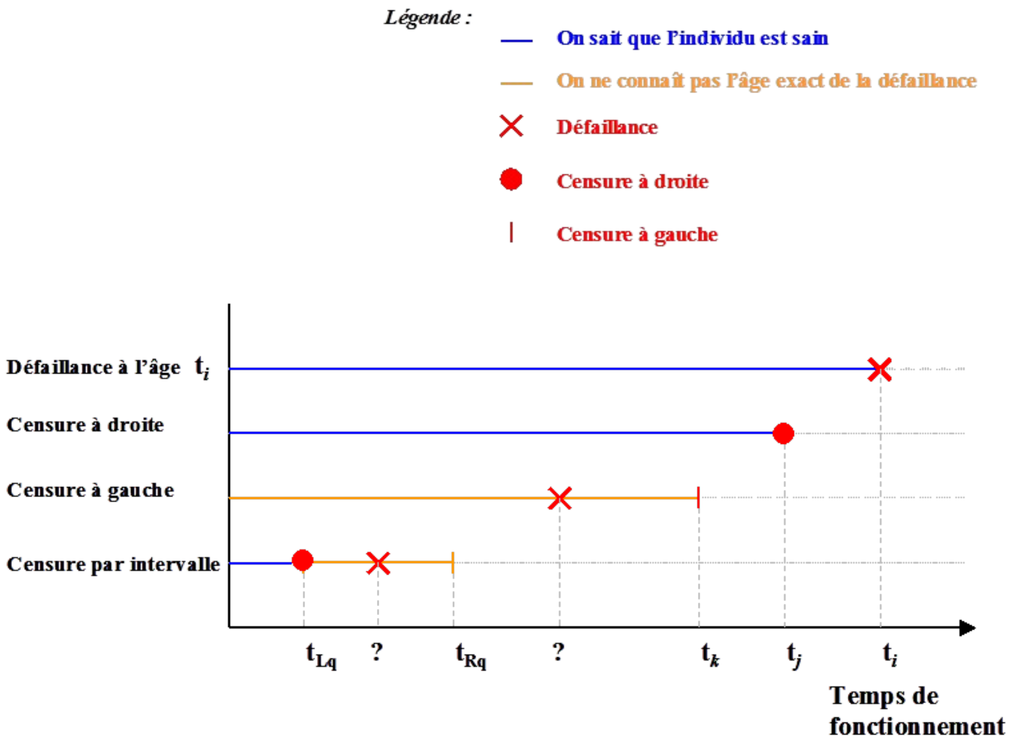
Défi: nous n'observons que des échantillons de l'ensemble de la population



Données censurées : savoir utiliser des information partielles

- Données censurées :
  - On dit qu'on a une observation censurée si la valeur de la mesure qui est faite n'est connue que partiellement.
  - Exemple d'application : Modélisation temps de défaillance d'une pièce dans une flotte.
    - On ne connait pas le temps de défaillance des pièces qui ne sont pas encore tombées en panne.
    - On sait seulement que ce temps de défaillance est supérieur à l'âge actuel de la pièce.

Trois types de censures
- À droite (suspension en anglais),
- À gauche,
- Par intervalle.

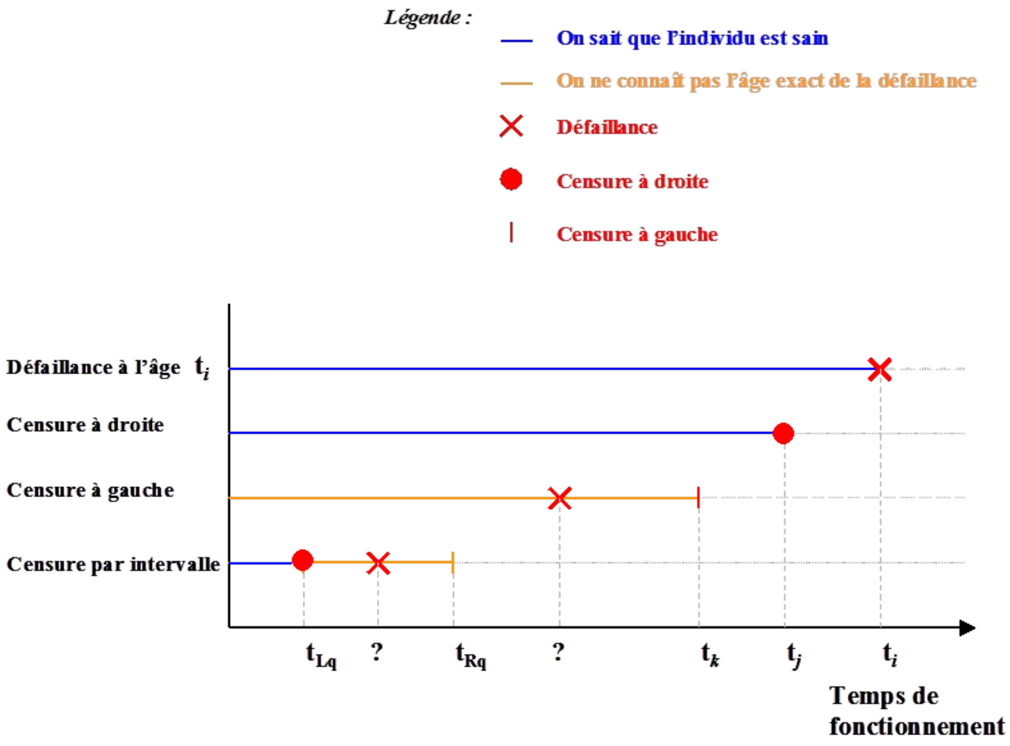


Données censurées, quelques notations

- Dans la suite, on travaille avec les censure à droite et on note :
  - Ti est le temps de défaillance, la date de censure
  - Yi = min(Ti, Ci) le temps observé
  - $\Delta i = I(Ti \leq Ci)$ un indicateur de censure à droite
- Pour avoir des résultats théoriques:
  - Hypothèse : les censures sont non informatives. C.à.d. la censure et le temps de défaillance sont indépendants :

$$T_i \perp \mathrm{C_i}$$

- Dans la réalité, la censure peut dépendre du temps.



Un peu de fiabilité, fonction de survie et taux de défaillance

- Soit T une variable aléatoire positive appelée temps de défaillance.
- $F(t) \coloneqq P(T \leq t)$ sa fonction de répartition (c.à.d. $F(t)$ est la probabilité d'avoir une défaillance avant le temps t)
- $f(t)$ est sa fonction de densité $f(t) = F'(t)$
- Fonction de survie : probabilité d'être encore en vie à temps t :

$$S(t) = P(T > t) = 1 - F(t)$$

- Taux de défaillance instantané : probabilité de tomber en panne entre t et t+dt sachant qu'on est en vie à temps t

$$h(t) = \lim_{dt \to 0} \frac{P(T < t + dt | T > t)}{dt} = \lim_{dt \to 0} \frac{F(t + dt) - F(t)}{dt} \times \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)}$$

- Taux de défaillance cumulé :

$$\mathrm{H}(t) = \int_0^t h(u)du = -\log\big(S(t)\big)$$

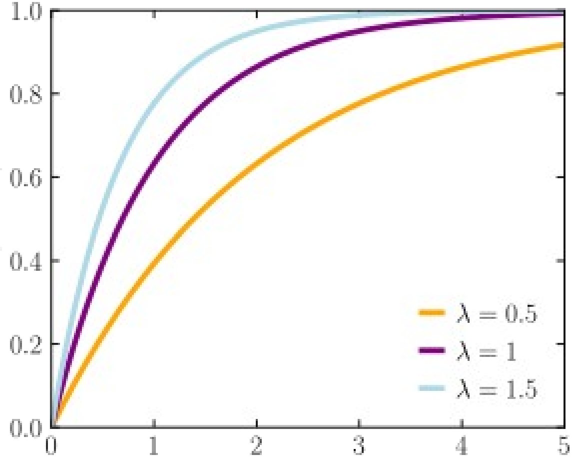
Un peu de fiabilité, fonction de survie et taux de défaillance

Expressed by	$S(t)$	$h(t)$
$S(t) =$	$-$	$\exp\left(-\int_0^t h(s)\,ds\right)$
$h(t) =$	$-\frac{d}{dt}\ln(S(t))$	$-$

Un peu de fiabilité, fonction de survie et taux de défaillance

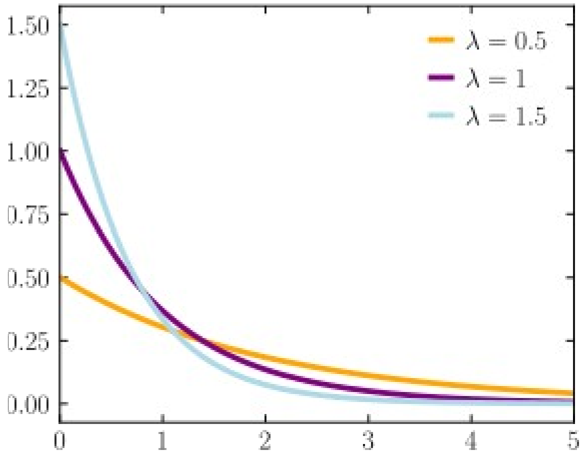
- Exemple simple : la loi exponentielle
  - > L'équipement vieilli continument <u>sans mémoire</u>

F(t)



t

f(t)



t

Un peu de fiabilité, les fonctions caractéristiques avec une loi exponentielle

- La fonction de répartition :

$$F(t) = \mathrm{P}(\mathrm{T} \leq \mathrm{t}) = 1 - \exp(-\lambda t)$$

- Fonction de survie :

$$S(t) = P(T > t) = \exp(-\lambda t)$$

- Taux de défaillance instantané :

$$h(t) = \frac{f(t)}{1 - F(t)} = \lambda$$

- Taux de défaillance cumulé :

$$H(t) = \int_0^t h(u)du = \lambda t$$

ESTIMATIONS PARAMÉTRIQUES

Maximum de vraisemblance

- Pour les méthodes paramétriques, on va chercher le maximum de vraisemblance pour un jeu de paramètres donné (on veut maximiser la vraisemblance des données)

- La vraisemblance dans la cas général est donnée par :

$$L\left(t_i, t_j, t_k, t_{Rq}, t_{Lq}; \theta\right) = \underbrace{\prod_{i=1}^{I} f_\theta(t_i)}\ \underbrace{\prod_{j=1}^{J}\left(1 - F_\theta(t_j)\right)}\ \underbrace{\prod_{k=1}^{K} F_\theta(t_k)}\ \underbrace{\prod_{q=1}^{Q}\left[F_\theta(t_{Rq}) - F_\theta(t_{Lq})\right]}$$

Défaillances ; « Probabilité instantanée » d'être défaillant à t_i

Censures à droite ; Probabilité d'être défaillant après t_j, sachant que j est sain à t_j

Censures à gauche ; Probabilité d'être défaillant avant t_k, sachant que k est défaillant à t_k

Censures par intervalle ; Probabilité d'être défaillant entre t_{Rq} et t_{Lq}, sachant que q est défaillant entre t_{Rq} et t_{Lq}

- On cherche ensuite le paramètre qui maximise la vraisemblance.
- Pour des raisons numériques, on maximise souvent la log-vraisemblance

Maximum de vraisemblance avec une loi exponentielle

- Application dans le cas d'une loi exponentielle avec des données censurées à droite ou complètes :
  - Expression de la loi :

$$f(t) = \lambda \cdot \exp(-\lambda t),\ F(t) = 1 - \exp(-\lambda t)$$

  - Echantillon (Y_i, δ_i) où $\delta_i = 1$ si défaillance, $\delta_i = 0$ sinon, on peut estimer λ directement :

$$\hat{\lambda} = \frac{\sum_{i=1}^{N} \delta_i}{\sum_{i=1}^{N} Y_i} = \frac{\text{Nombre de défaillances}}{\text{Temps cumulé de fonctionnement}} = \frac{d}{T_{cum}}$$

Maximum de vraisemblance avec une loi exponentielle

- Application dans le cas d'une loi exponentielle avec des données censurées à droite ou complètes :
  - Expression de la loi : $f(t) = \lambda \cdot \exp(-\lambda t)$, $F(t) = 1 - \exp(-\lambda t)$
  - Echantillon (Y_i, δ_i) où $\delta_i = 1$ si défaillance, $\delta_i = 0$ sinon, on peut estimer λ directement :

$$\hat{\lambda} = \frac{\sum_{i=1}^{N} \delta_i}{\sum_{i=1}^{N} Y_i} = \frac{\text{Nombre de défaillances}}{\text{Temps cumulé de fonctionnement}} = \frac{d}{T_{cum}}$$

- **En effet →**

$$L(t_i, t_j, t_k, t_{Rq}, t_{Lq}; \theta) = \prod_{i=1}^{I} f_\theta(t_i) \prod_{j=1}^{J} (1 - F_\theta(t_j)) \prod_{k=1}^{K} F_\theta(t_k) \prod_{q=1}^{Q} \left[F_\theta(t_{Rq}) - F_\theta(t_{Lq})\right]$$

Maximum de vraisemblance avec une loi exponentielle

- Application dans le cas d'une loi exponentielle avec des données censurées à droite ou complètes :
  - Expression de la loi : $f(t) = \lambda \cdot \exp(-\lambda t)$, $F(t) = 1 - \exp(-\lambda t)$
  - Echantillon (Y_i, δ_i) où $\delta_i = 1$ si défaillance, $\delta_i = 0$ sinon, on peut estimer λ directement :

$$\hat{\lambda} = \frac{\sum_{i=1}^{N} \delta_i}{\sum_{i=1}^{N} T_i} = \frac{\text{Nombre de défaillances}}{\text{Temps cumulé de fonctionnement}} = \frac{d}{T_{cum}}$$

- **En effet →**

$$L(t_i, t_j, t_k, t_{Rq}, t_{Lq}; \theta) = \boxed{\prod_{i=1}^{I} f_\theta(t_i) \prod_{j=1}^{J} (1 - F_\theta(t_j))}$$

$$L(T_i; \lambda) = \prod_i f(T_i)^{\delta_i} \times (1 - F(T_i))^{1-\delta_i} = \prod_{i,\delta_i=1} \lambda \cdot \exp(-\lambda T_i) \cdot \prod_{i,\delta_i=0} \exp(-\lambda T_i) = \lambda^d \cdot \exp(-\lambda T_{cum})$$

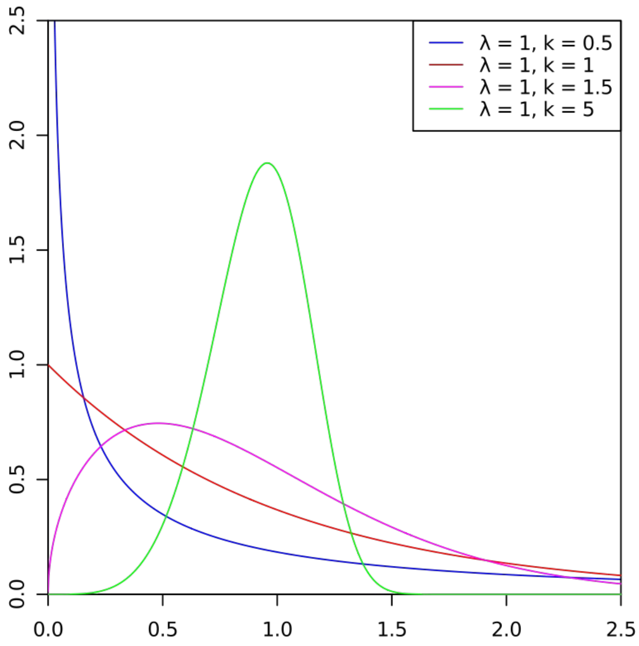
$\frac{\partial L}{\partial \lambda} = -d\lambda^{d-1} \cdot \exp(-\lambda T_{cum}) + T_{cum}\lambda^d \cdot \exp(-\lambda T_{cum})$ ou en log : $\frac{\partial \log(L)}{\partial \lambda} = \frac{d}{\lambda} - T_{cum}$

$$\frac{\partial^2 \log(L)}{\partial \lambda^2} = \frac{d}{\lambda^2} > 0$$

Maximum de vraisemblance avec une loi de Weibull

- **Application dans le cas d'une loi de Weibull avec des données complètes :**
  - Le modèle exponentiel est souvent trop simple

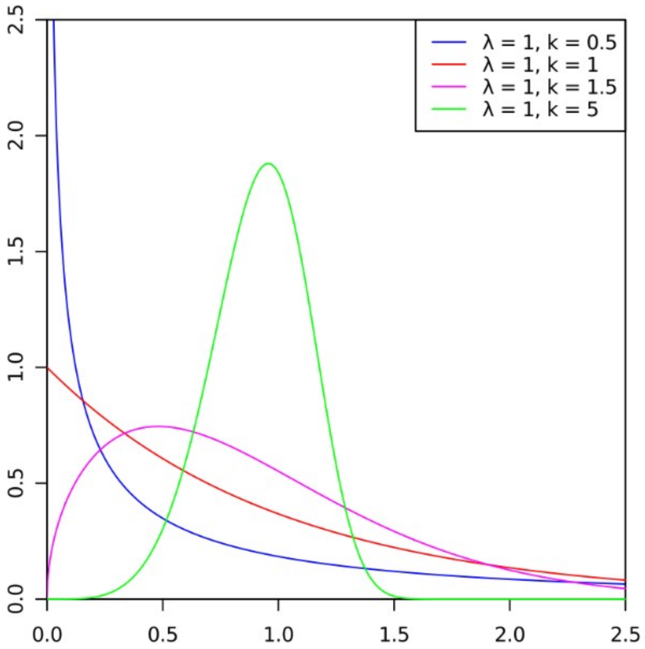
- **Modèle de Weibull**
  - $f(t,\lambda,k) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k} e^{\left(-\frac{t}{\lambda}\right)^{k}}$
  - $k < 1$: le taux de panne diminue avec le temps
  - $k = 1$: le taux de panne constant dans le temps
  - $k > 1$: le taux de panne augmente avec le temps
  - k est le paramètre de forme
  - λ est le parameter d'échelle



Maximum de vraisemblance avec une loi de Weibull

- **Détermination des coefficients**
  - > Méthode “graphique”
    - Calcul des probabilités cumulées de pannes F(t)
    - On trace log(t) = log(-log(1-F(t)))
    - On a :
      - $F(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}$
      - $-\ln\big(1 - F(x)\big) = \left(\frac{x}{\lambda}\right)^k$
      - $\ln\big(-\ln\big(1 - F(x)\big)\big) = kln(x) - kln\lambda$
    - On en déduit facilement k et λ

  - > Méthode par maximum de vraisemblance
    - Connaissant k:
      - $\lambda^k = \frac{1}{n}\sum_i t_i^k$
    - Pour k:
      - $0 = \frac{\sum_i t_i^k \ln(t_i)}{\sum_i t_i^k} - \frac{1}{k} - \frac{1}{n}\sum_i \ln(t_i)$
      - Généralement, cette equation est résolue numériquement



ESTIMATIONS NON PARAMÉTRIQUES

Estimation non paramétrique

- Objectif : estimer de manière non paramétrique des fonctions représentatives de la distribution de nos défaillances à partir d'un échantillon de données censurées

Estimation non paramétrique : Kaplan-Meier

Samples	I	II
Initial numbers	100	1000
Deaths in first year of age	70	750
One-year survivors	30	250
Deaths in second year of age	15	
Two-year survivors	15	

- Quelle est la probabilité de survivre 1 an ?

Estimation non paramétrique : Kaplan-Meier

Samples	I	II
Initial numbers	100	1000
Deaths in first year of age	70	750
One-year survivors	30	250
Deaths in second year of age	15	
Two-year survivors	15	

- Quelle est la probabilité de survivre 1 an ?

$$\hat{P}(1) = P^*(1) = (30 + 250)/(100 + 1000) = .255$$

Estimation non paramétrique : Kaplan-Meier

Samples	I	II
Initial numbers	100	1000
Deaths in first year of age	70	750
One-year survivors	30	250
Deaths in second year of age	15	
Two-year survivors	15	

- Quelle est la probabilité de survivre 1 an ?

$$\hat{P}(1) = P^*(1) = (30 + 250)/(100 + 1000) = .255$$

- Deux ans ?

Estimation non paramétrique : Kaplan-Meier

Samples	I	II
Initial numbers	100	1000
Deaths in first year of age	70	750
One-year survivors	30	250
Deaths in second year of age	15	
Two-year survivors	15	

$$\hat{P}(1) = P^*(1) = (30 + 250)/(100 + 1000) = .255$$

$$\hat{P}(2)/\hat{P}(1) = 15/30 = 0.50,$$

$$\hat{P}(2) = 0.255 \times 0.50 = 0.127,$$

Estimation non paramétrique : Kaplan-Meier

$t_2 > t_1$: deux temps de défaillance successifs

$$\tilde{S}(t_2) = P(T > t_2) = P(T > t_2 \wedge T > t_1)$$

$$= P(T > t_2 | T > t_1) \times P(T > t_1) \quad = \left(1 - \frac{d_2}{n_2}\right) \times S(t_1)$$

- d_2 : nombre de défaillances dans le temps t_2
- n_2: nombre d'individus à risque à t_2, (c.à.d. k tels que $y_k \geq t_2$)

Estimation non paramétrique : Kaplan-Meier

- Estimateur de Kaplan-Meier : estimateur non paramétrique de la fonction de survie
  - $\tilde{S}(t) = \prod_{t_i < t}\left(1 - \frac{d_i}{n_i}\right)$
  - d_i : nombre de défaillances dans le temps t_i
  - n_i: nombre d'individus à risque dans le temps t_i, (c.à.d. k tels que $y_k \geq t_i$)

- Estimateur de Kaplan-Meier = Estimateur de maximum de vraisemblance sous hypothèses de régularités → on peut construire la variance de l'estimateur et des intervalles de confiance

Estimation non paramétrique : Kaplan-Meier

- Example:

	Indicateur de défaillance
6	1
19	1
32	1
42	1
42	1
43	0
94	1
126	0
169	0
207	1

	Indicateur de défaillance
211	0
227	0
253	1
255	0
270	0
310	0
316	0
335	0
346	0

Source exemple : Gilbert Colletaz, note de cours

Estimation non paramétrique : Kaplan-Meier

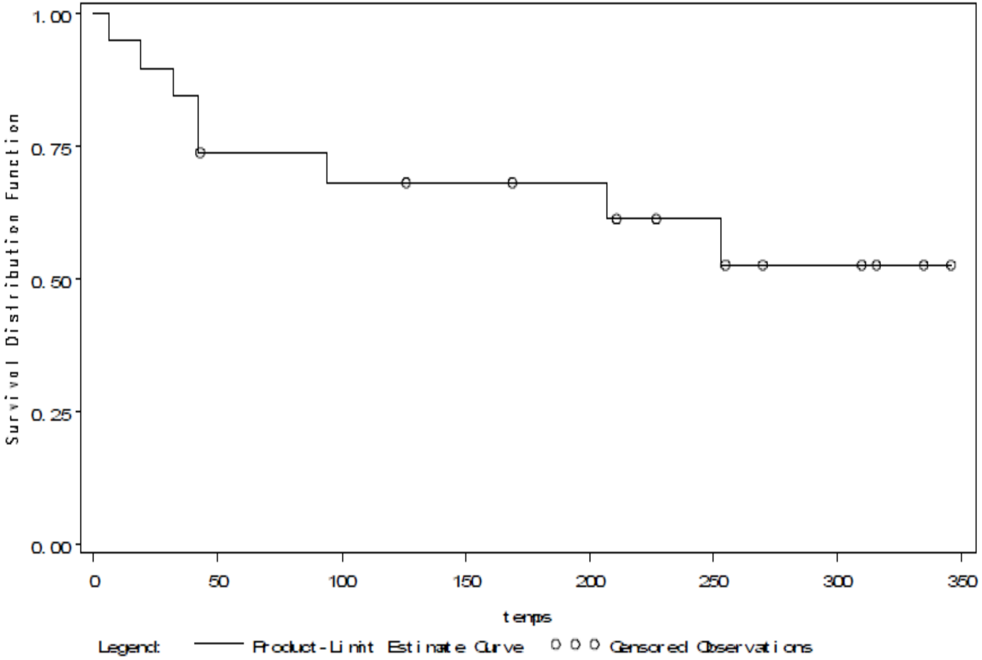
6, 19, 32, 42, 42, 43*, 94, 126*, 169*, 207, 211*, 227*, 253, 255*, 270*, 310*, 316*, 335*, 346*

On obtient alors :

t_i	d_i	n_i	$1 - d_i/n_i$	$\hat{S}(t_i)$	$\hat{F}(t_i) = 1 - \hat{S}(t_i)$
0	0	19	1	1	0
6	1	19	0.947	0.947	0.053
19	1	18	0.944	0.895	0.105
32	1	17	0.941	0.842	0.158
42	2	16	0.875	0.737	0.263
94	1	13	0.923	0.680	0.320
207	1	10	0.90	0.612	0.388
253	1	7	0.957	0.525	0.475

Estimation non paramétrique : Kaplan-Meier

6, 19, 32, 42, 42, 43*, 94, 126*, 169*, 207, 211*, 227*, 253, 255*, 270*, 310*, 316*, 335*, 346*



Estimation non paramétrique : Nelson-Aalen

- Estimateur de Nelson-Aalen : estimateur non paramétrique du taux de défaillance cumulé.
  - $\widetilde{H}(t) = \sum_{t_i < t} \frac{d_i}{n_i}$
  - d_i : Nombre de défaillances dans t_i
  - n_i : Nombre d'individus à risque dans t_i

- Intuition :
  - d_i/n_i est un estimateur du taux de défaillance instantané

Estimation non paramétrique : Autres

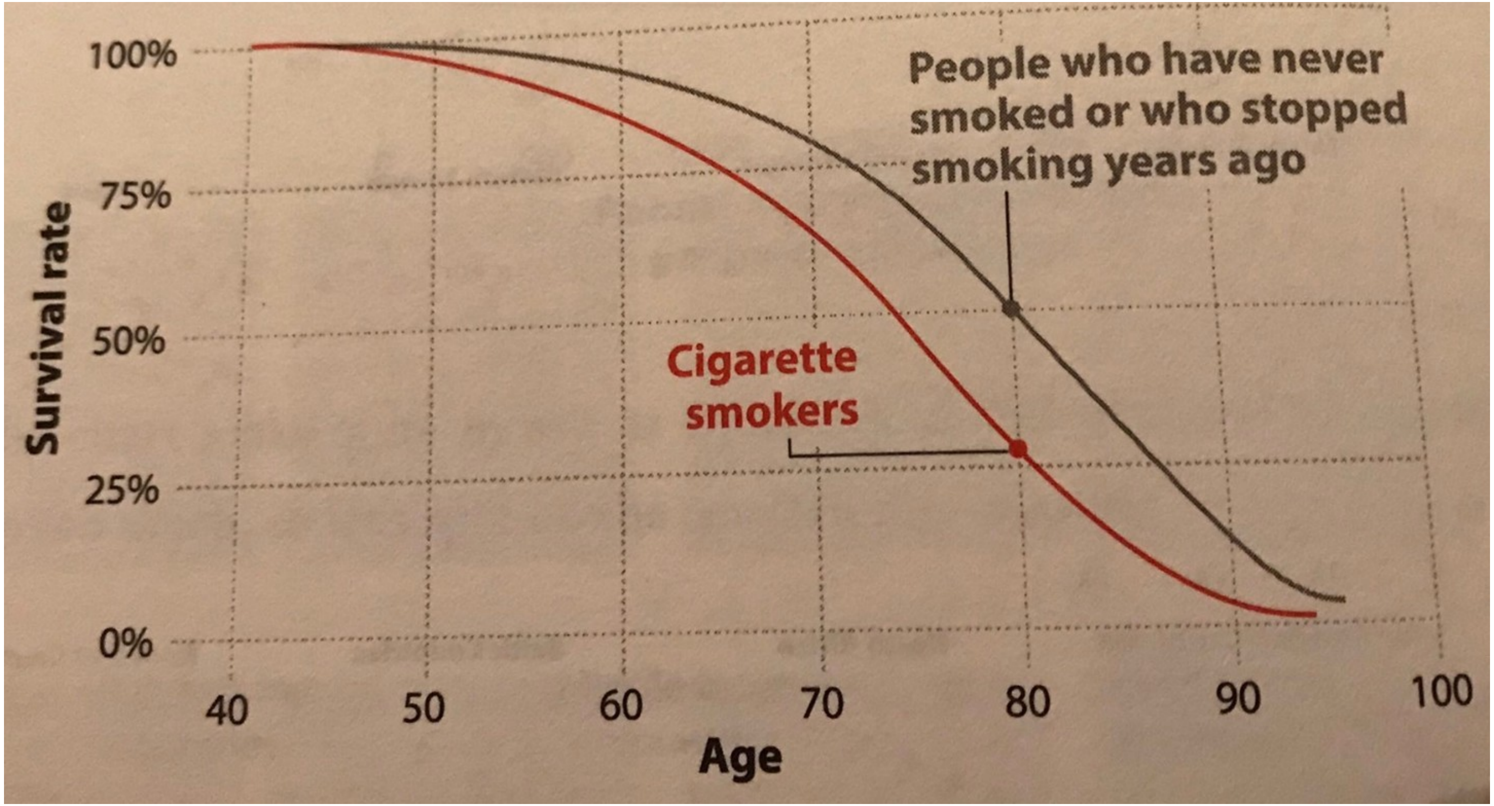
- D'autres estimateurs :
  - Méthode de Johnson sur les rangs médians ou moyen
  - Estimateur de Fleming-Harrington
  - Estimateur de Breslow…

INFLUENCES

Influences

- Est-ce raisonnable de considérer que des systèmes produits de manière identique ont tous la même probabilité de défaillance ?

Influences



100%
75%
50%
25%
0%
Survival rate
People who have never smoked or who stopped smoking years ago
Cigarette smokers
40
50
60
70
80
90
100
Age

Influences

Certains modèles incorporent des covariates/features

- Cox proportional hazards model
- Aalen’s additive model
- Weibull Accelerated Failure time
- Random Survival Forest
- Gradient boosting Cox log partial likelihood loss
- DeepSurv

Cox Proportional Hazards Model

- **Les covariates**
  - Features qui représentent la présence ou l'absence d'un phénomène
  - Ex: fumer / ne pas fumer faire du sport / ne pas faire du sport, etc…
- **Equation de la loi de vie :**
  - $\lambda(t, X_0, \dots, X_n) = \lambda_0(t)\exp(\sum_{i=1..n} \beta_i X_i)$ (taux de défaillance instantanée)
  - λ_0
    - le **risque de base**
    - Il dépend du temps
    - il est généralement non estimé (car on veut étudier le rapport entre les variables) → semi-paramétrique
  - $\exp(\sum_{i=1..n} \beta_i X_i)$
    - Dépend uniquement des covariates
    - Ne dépend pas du temps (modèle de Cox simple)

- **Risques proportionnels → comparaison de deux individus**
  - Avec 2 individus qui varient uniquement sur le covariate k:
    - $\frac{\lambda(t,I_1)}{\lambda(t,I_2)} = \frac{\lambda_0(t)\ (\exp(\sum_{i\neq k} \beta_i X_i + 1.\beta_k))}{\lambda_0(t)\ (\exp(\sum_{i\neq k} \beta_i X_i + 0.\beta_k))} = \exp(\beta_k)$
- **Les X font varier les risques**

Accelerated Failure Time

- **Principe**
  > Hypothèse : les facteurs influents accélèrent ou ralentissent le temps avant événement

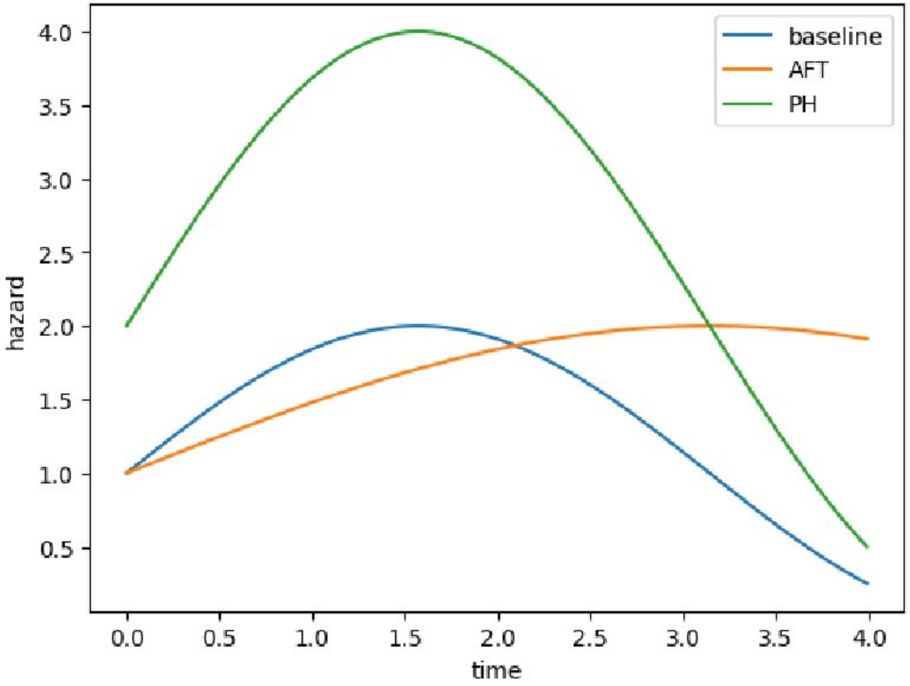
  > Le modèle devient donc :
  - $\lambda(t|X) = \exp(-\beta x)\,\lambda_0(\exp(-\beta x)\,t)$
  - Soit $S(t|X) = S_0(\exp(-\beta x)\,t)$, avec S_0 une fonction de survie quelconque

  > On peut aussi l'écrire ainsi (T le temps de survie):
  - $\log(T) = \beta x + \epsilon$
  - ϵ suivra une distribution qui changera la forme de la fonction de survie de T

  > Les hypothèses :
  - On connait S_0
  - Les X_i ont une interaction linéaire et indépendante

  > Ici, on agit sur le temps et pas sur le risque

Accelerated Failure Time vs Proportional Hazard



DeepSurv

- **DeepSurv est une amélioration du modèle de Cox**
  - > $\lambda(t, X_0, \dots, X_n) = \lambda_0(t)\exp(\sum_{i=1..n} \beta_i X_i)$ devient :
  - > $\lambda(t, X_0, \dots, X_n) = \lambda_0(t)\exp(h_\theta(X_i))$
  - > Où h_θ est un réseau de neurones
  - > La fonction de coût est adaptée :

$$l(\theta) := -\sum_{i:E_i=1} \left(\hat{h}_\theta(x_i) - \log \sum_{j \in \Re(T_i)} e^{\hat{h}_\theta(x_j)} \right).$$

  - > Comparaison de deux traitements différents :

$$\mathrm{rec}_{ij}(x) = \log\left(\frac{\lambda(t; x|\tau = i)}{\lambda(t; x|\tau = j)}\right) = \log\left(\frac{\lambda_0(t) \cdot e^{h_i(x)}}{\lambda_0(t) \cdot e^{h_j(x)}}\right) = h_i(x) - h_j(x).$$

Modèle additif d’Aalen

- **Principe**
  - Le modèle de Cox fait l’hypothèse que l’influence des covariables ne change pas selon le temps
  - Le modèle additif d’Aalen fait une proposition différente :

$$\lambda(t) = (\beta_0) + \sum \beta_i(t) x_i$$

  - Soit N(t) le nombre d’événement jusqu’à t
  - On doit introduire l’indicateur de risque δ_i, tel que $\delta_i = 1\ si\ X_i\ est\ à\ risque, 0\ sinon$ (comme pour Kaplan-Meier)
  - On pose $X(t) = (\delta_0 X_0(t), \dots, \delta_n X_n(t))$
  - On a alors : $\lambda(t) = X'(t)\beta(t)$
  - On a $\Delta N(t) = \Delta M(t) + \lambda(t)$: la différence des états entre t et t-1 dépend d’une certaine quantité M et de λ.
  - Ici je passe la quantité ΔM, (qui veut peut regarder Shrinkage Estimation for Aalen's Additive Model (uwindsor.ca)), mais il se trouve que $E(M) = 0$, donc on peut utiliser une regression aux moindres carrés pour trouver les β, qui vont valoir :

$$\beta(t) = X^{-1}(t)\Delta \mathrm{N}(t)$$

Random Survival Forest

- **Principe**
  - > Même chose que les Random Forest classiques, mais avec des « survival tree »
  - > Le survival tree :
    - Couper selon les covariates (comme les arbres, ça peut être sur une coordonnée, plusieurs, toutes, etc..)
    - Critère de découpe
      - Le plus courant : C-index
      - Il en existe d'autres : L1 splitting, AUC splitting, Maximally selected rank splitting

- **Le Concordance-Index:**
  - > $C - index = \frac{\sum_{i,j} I(t_i > t_j) I(\hat{t_i} > \hat{t_j}) \delta_j}{\sum_{i,j} I(t_i > t_j) \delta_j}$
  - > δ_j vaut 0 si on la variable est censurée

WANG, Hong et LI, Gang. A selective review on random survival forests for high dimensional data. *Quantitative bio-science*, 2017, vol. 36, no 2, p. 85.

LES MÉTRIQUES

Mesurer la performance des méthodes

- **Comment mesurer la performance de nos algorithmes, alors que la vérité terrain est censurée ?**
  - On a l'habitude d'utiliser la RMSE (Root Mean Square Error)
  - On a prédit/modéliser des distributions, et non des prédictions uniques pour chaque sample
  - De plus, quand les données sont censurées, cette mesure est impossible sans adaptation !

- **Exemple simple:**
  - La loi de défaillance suit une loi de Weibull
  - On apprend la loi
  - Comment faire une prédiction ?
    - Moyenne ?
    - Médiane ?
  - Dans tous les cas, on aura un résidu incompressible, même si notre loi apprise correspond parfaitement au phénomène sous-jacent
  - Enfin, même en prédisant la médiane, dans le cas où la vérité terrain est censurée, on ne sait pas comment calculer la RMSE avec un point inconnu

  - Il faut donc raisonner sur tout le dataset, et non sample par sample

Mesurer la performance des méthodes

- **Le Concordance-Index (qu'on a vu dans les slides précédents):**
  - > $C_{index} = \frac{\sum_{i,j} I(t_i > t_j) I(\hat{t_i} > \hat{t_j}) \delta_j}{\sum_{i,j} I(t_i > t_j) \delta_j}$
  - > δ_j vaut 0 si on la variable est censurée

  - > Le principe est de classer au mieux les paires
  - > 1 est un score parfait
  - > 0,5 signifie qu'on ne classe pas mieux que le hasard

  - > Le défaut évident :
    - Si on fait d'énormes erreurs, mais que les échantillons sont bien classés, on a un score de 1

Mesurer la performance des méthodes

- **Le Brier Score**
  - Prévu pour comparer des distributions par rapport à des tirages d'événements
  - Le score : $\frac{1}{n}\sum_{i=0}^{n-1}(o_i - f_i)^2$
  - où :
    - n est le nombre d'événements
    - f_i est la prévision (une probabilité de 0 à 1) pour le i-ème événement
    - o_i est le résultat (0 ou 1) du i-ème événement
  - Exemple :
    - Comment comparer deux modèles de vainqueur pour la présidentielle ?
    - Une fois l'élection passée, on compare le pourcentage obtenu avec le résultat, le modèle le plus proche a un meilleur score

Mesurer la performance des méthodes

- **Le Brier Score temporel**
  - $BS(t) := \frac{1}{n}\sum_{i=0}^{n-1}\left(I(y_i > t) - \hat{S}(t|x_i)\right)^2$
  - où :
    - n est le nombre d'événements
    - $I(\cdot)$ est la function indicatrice
    - $\hat{S}(t|x)$ est la probabilité prédite d'un modèle de rester sans événement jusqu'à l'instant t pour le vecteur de caractéristiques x
  - Exemple :
    - Mon modèle donne une probabilité pour chaque temps que l'événement arrive, je l'évalue en t
    - On compare cela au fait que l'événement ait eu lieu $I(y > t)$

Mesurer la performance des méthodes

- **Le Brier Score temporel avec données censurées**
  - $BS(t) := \frac{1}{n}\sum_{i=0}^{n-1}\left(I(y_i > t) - \hat{S}(t|x_i)\right)^2$ peut se réécrire

$$BS(t) := \frac{1}{n}\sum_{i=0}^{n-1} I(y_i \leq t)\left(0 - \hat{S}(t|x_i)\right)^2 + I(y_i > t)\left(1 - \hat{S}(t|x_i)\right)^2$$

  - où :
    - n est le nombre d'événements
    - $I(\cdot)$ est la function indicatrice
    - $\hat{S}(t|x)$ est la probabilité prédite d'un modèle de rester sans événement jusqu'à l'instant t pour le vecteur de caractéristiques x (i.e. probabilité d'être encore en vie)

Mesurer la performance des méthodes

- **Le Brier Score temporel avec données censurées**
  - $BS(t) := \frac{1}{n}\sum_{i=0}^{n-1}\left(I(y_i > t) - \hat{S}(t|x_i)\right)^2$ peut se réécrire $BS(t) := \frac{1}{n}\sum_{i=0}^{n-1} I(y_i \leq t)\left(0 - \hat{S}(t|x_i)\right)^2 + I(y_i > t)\left(1 - \hat{S}(t|x_i)\right)^2$
  - On va utiliser cette formalisation pour faire intervenir la censure
  - Soit $\hat{G}(t) \coloneqq P(C > t)$ l'estimateur de la fonction de survie conditionnelle des temps de censure calculés selon la méthode de Kaplan-Meier, où C est le temps de censure (et non l'événement!)
  - $BS^c(t) := \frac{1}{n}\sum_{i=0}^{n-1} I(y_i \leq t \wedge \delta_i = 1)\frac{\left(0-\hat{S}(t|x_i)\right)^2}{\hat{G}(y_i)} + I(y_i > t)\frac{\left(1-\hat{S}(t|x_i)\right)^2}{\hat{G}(t)}$,
  - où :
    - n est le nombre d'événements
    - $I(\cdot)$ est la function indicatrice
    - $\hat{S}(t|x)$ est la probabilité prédite d'un modèle de rester sans événement jusqu'à l'instant t pour le vecteur de caractéristiques x (i.e. probabilité d'être encore en vie)
    - $1/\hat{G}(\cdot)$ est un poids de pondération qui utilise l'estimation de la probabilité de censure : on peut voir cette partie comme une pondération par l'inverse de la probabilité de censure pour ce temps -> grande probabilité d'être censuré = peu de fiabilité sur l'erreur, petite probabilité d'être censuré -> grande fiabilité de l'erreur

Mesurer la performance des méthodes

- **Le Brier Score temporel avec données censurées intégré**
  - Une fois le Brier Score calculé, on va tout simplement l'intégrer sur tous les temps.
  - Ce score est calculé par une approximation trapézoïdale (que ce soit censuré ou non censuré)
  - $IBS = \int_{t_1}^{t_{max}} BS(t)\, dw(t)$

Source : pysurvival documentation

CONCLUSIONS

Conclusions

- **Méthode statistiques**
  - > On veut modéliser les probabilités de pannes pour un système
  - > On dispose d'échantillons pour le faire
  - > Ces échantillons peuvent être censurés (le plus souvent à droite)

- **Les outils pour le faire**
  - > Loi paramétriques
  - > Loi non paramétriques

- **Influences**
  - > Modèles spécifiques
  - > On peut vouloir uniquement l'analyse relative
  - > Quelques exemples emblématiques ont été passés en revue

- **Métriques**
  - > Concordance Index
  - > Brier Scores