Cours 1 : notes

Slide (8)

$$\lambda(t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}(t \leq T \leq t+h \,/\, T \geq t)$$

$$= \lim_{h \downarrow 0} \frac{1}{h} \frac{\mathbb{P}(t \leq T \leq t+h , T \geq t)}{\mathbb{P}(T \geq t)}$$

$$= \lim_{h \downarrow 0} \frac{1}{h} \frac{\mathbb{P}(t \leq T \leq t+h)}{\mathbb{P}(T \geq t)} \longrightarrow f(t)$$

$$= \frac{f(t)}{\bar{F}(t)}$$

$\bar{F}(t) = \mathbb{P}(T > t)$

mais ici comme T a une loi continue

$\mathbb{P}(T > t) = \mathbb{P}(T \geq t)$

Slide (9). $T \sim W(\lambda, \alpha)$ densité $\alpha \lambda^\alpha t^{\alpha - 1} \exp(-(\lambda t)^\alpha)$ sur $\mathbb{R}_+$

fdr (fonction de répartition)

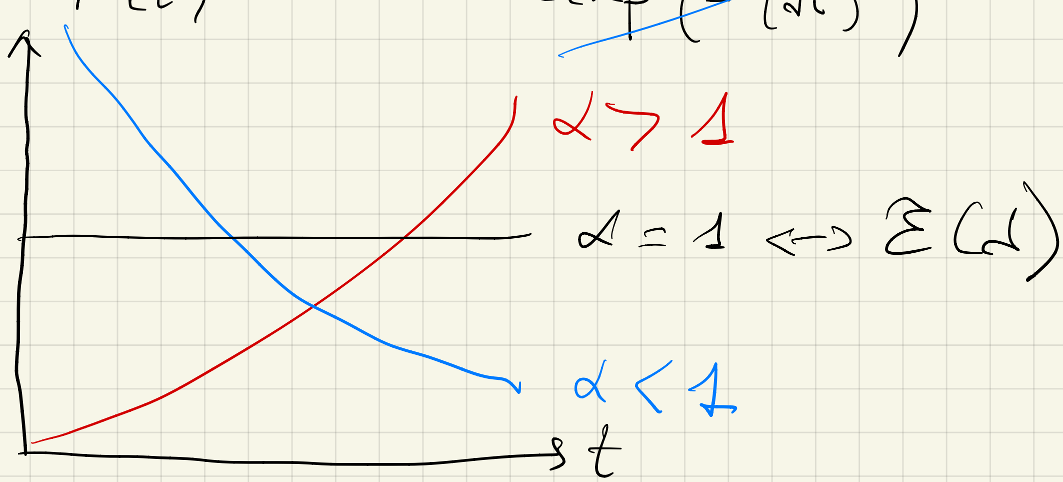
$F(t) = 1 - e^{-(\lambda t)^\alpha}$ sur $\mathbb{R}_+$

fonction de survie

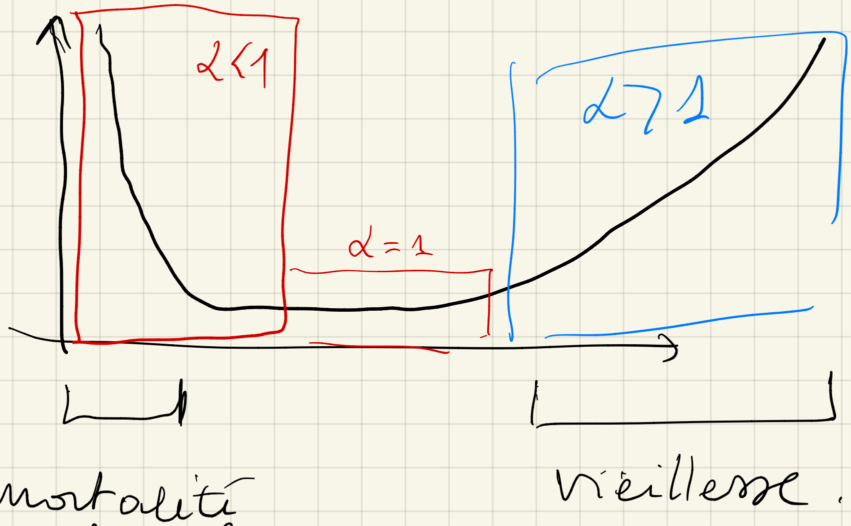
$\bar{F}(t) = 1 - F(t) = e^{-(\lambda t)^\alpha}$ $\quad \underset{t=0}{1} \longrightarrow \underset{t=\infty}{0}$

hazard rate

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)} = \frac{\alpha \lambda^\alpha t^{\alpha-1} \cancel{\exp(-(\lambda t)^\alpha)}}{\cancel{\exp(-(\lambda t)^\alpha)}}$$



Remarque : hazard rate pour une pop. humaine

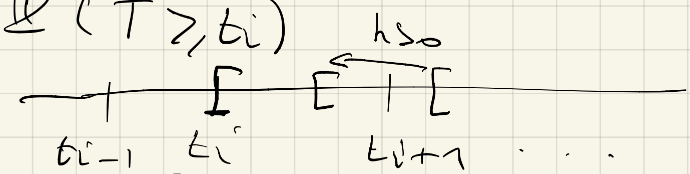


cumulative hazard $\Lambda(t) = \int_0^t \lambda(x)\,dx$

$= \int_0^t \alpha \lambda^\alpha x^{\alpha-1}\,dx$

Slide (10)

$\lambda(t_i) = \lim_{h \searrow 0} \dfrac{\mathbb{P}(t_i \leq T \leq t_i+h)}{\mathbb{P}(T \geq t_i)}$



$= \dfrac{\mathbb{P}(T = t_i)}{\mathbb{P}(T \geq t_i)} = p_i$

⚠ $\mathbb{P}(T \geq t_i) = \mathbb{P}(T > t_{i-1}) = \bar{F}(t_{i-1})$

$= p_i + p_{i+1} + \dots$

Slide (11).

Exercice $0 < t_1 \le t_2 \le t_3 \le \dots$ $\mathbb{P}(T = t_i) = p_i$

$$\prod_{j=1}^{i}\left(1 - \lambda(t_j)\right) = \prod_{j=1}^{i}\left(1 - \frac{p_j}{\sum_{k:t_k\ge t_j} p_k}\right)$$

$$= \left(1-\frac{p_1}{\sum_{k:t_k\ge t_1}p_k}\right)(*)\left(1-\frac{p_2}{\sum_{k:t_k\ge t_2}p_k}\right)(**)\cdots\cdots\left(1-\frac{p_i}{\sum_{k:t_k\ge t_i}p_k}\right)(***)$$

$(*)$ $\displaystyle\sum_{k:t_k\ge t_1} p_k = p_1+p_2+p_3+\dots = 1$

$(**)$ $\displaystyle\sum_{k:t_k\ge t_2} p_k = p_2+p_3+\dots$

$(***)$ $\displaystyle\sum_{k:t_k\ge t_i} p_k = p_i+p_{i+1}+\dots$

$$= \left(1-\frac{p_1}{1}\right)\left(1-\frac{p_2}{p_2+p_3+\dots}\right)\left(1-\frac{p_3}{p_3+p_4+\dots}\right)\cdots\left(1-\frac{p_i}{p_i+p_{i+1}+\dots}\right)$$

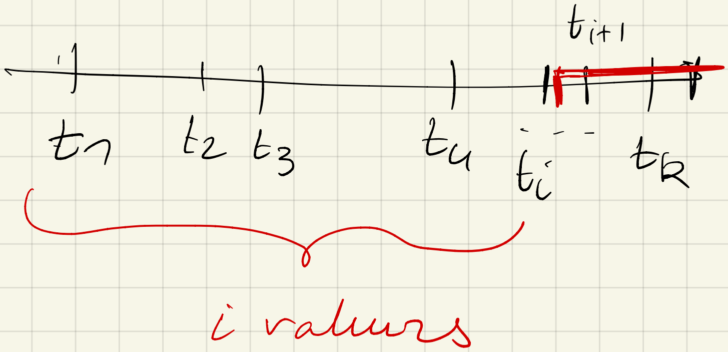
$$= \cancel{(1-p_1)}\left(\frac{\cancel{p_2}+p_3+\dots-\cancel{p_2}}{\cancel{p_2+p_3+\dots}}\right)\left(\frac{p_4+\cancel{p_3}+\dots}{\cancel{p_3+p_4+\dots}}\right)\cdots\left(\frac{p_{i+1}+p_{i+2}+\dots}{\cancel{p_i+p_{i+1}+\dots}}\right)$$

On a $p_1+p_2+p_3+\dots = 1 \iff p_2+p_3+\dots = 1-p_1$

$$= p_{i+1}+p_{i+2}+\dots = \mathbb{P}(T\ge t_{i+1}) = \mathbb{P}(T>t_i)$$
$$= \overline{F}(t_i).$$

Exercice 2 $\quad T \sim \mathcal{U}\{t_1 \leq t_2 \leq \dots \leq t_k\}$ $\quad \mathbb{P}(T = t_i) = \frac{1}{k}$.

Fonction de survie $\quad \bar{F}(t_i) = \mathbb{P}(T > t_i)$



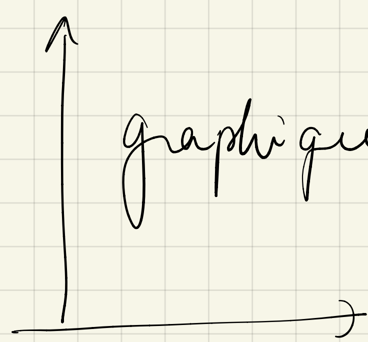
$$= \frac{k - i}{k} = 1 - \frac{i}{k}$$

Hazard rate $\quad \lambda(t_i) = \dfrac{\mathbb{P}(T = t_i)}{\bar{F}(t_{i-1})} = \dfrac{\frac{1}{k}}{1 - \frac{i-1}{k}} = \dfrac{1}{k - i + 1}$.

Slide (16)

Patient	Obs. time (T^C)	Status (δ)
1	7	0
2	6	1
3	6	0
4	5	0
5	2	1
6	4	1

$\Leftarrow$



Machine learning : label (T^C, δ).
$$\underset{\mathbb{R}_+}{\Uparrow} \quad \underset{\{0,1\}}{\Uparrow}$$

de plus on veut la loi de T.

Slide (18)

$$\delta = \mathbb{1}_{(T \le C)}$$

$$\frac{d}{dt}\mathbb{P}(T^c \le t, \delta = 1) \overset{?}{=} f(t)\,\bar{G}(t)$$

$$\mathbb{P}(T^c \le t, \delta = 1) = \mathbb{P}(T \le t, T \le C)$$

T densité f
C densité g
et $\underbrace{T \perp\!\!\!\perp C}$.

$$= \iint \mathbb{1}(u \le t, u \le v)\, f(u)\, g(v)\, du\, dv$$

$$= \int \mathbb{1}_{(u \le t)} f(u) \underbrace{\int_u^{\infty} g(v)\, dv}_{\bar{G}(u)}\, du$$

$$G(u) = \int_{-\infty}^{u} g(v)\, dv.$$

$$= \int_0^t f(u)\, \bar{G}(u)\, du.$$

$$\frac{d}{dt}\mathbb{P}(T^c \le t, \delta = 1) = f(t)\,\bar{G}(t).$$

$$\mathbb{P}(T^c \le t, \delta = 0) = \mathbb{P}(C \le t, C \le T)$$
$$= \dots = \int_0^t g(v)\,\bar{F}(v)\, dv.$$

Slide (19)

Data $(T_1^c, \delta_1) \dots (T_n^c, \delta_n)$ iid.

"la densité de $(T^c, \delta = 1)$ est $f(t)\bar{G}(t)$"
" ———— $(T^c, \delta = 0)$ est $g(t)\bar{F}(t)$"

$$\mathcal{L}\left((T_1^c, \delta_1), \dots, (T_n^c, \delta_n)\right) = \prod_{i=1}^{n} \left(f(T_i^c)\,\bar{G}(T_i^c)\right)^{\delta_i} \left(g(T_i^c)\,\bar{F}(T_i^c)\right)^{1-\delta_i}$$

$$= \boxed{\underbrace{\prod_{i=1}^{n} f(T_i^c)^{\delta_i}\, \bar{F}(T_i^c)^{1-\delta_i}}_{\text{dépend de } f, \bar{F} \text{ de la loi de } T}} \underbrace{\prod_{i=1}^{n} g(T_i^c)^{1-\delta_i}\, \bar{G}(T_i^c)^{\delta_i}}_{\text{dépend de la loi de } C}$$

Slide 20

$T \sim \mathcal{E}(\lambda)$

Data $(T_1^c, \delta_1), \dots, (T_n^c, \delta_n)$

But : estimer λ.

$f(t) = \lambda e^{-\lambda t}$

$\bar{F}(t) = e^{-\lambda t}$.

$\mathbb{E}(T) = \frac{1}{\lambda}$

Part de la vraisemblance qui dépend de la loi de T

$$\prod_{i=1}^n f(T_i^c)^{\delta_i} \bar{F}(T_i^c)^{1-\delta_i}$$

$$= \prod_{i=1}^n \left(\lambda \underbrace{e^{-\lambda T_i^c}}\right)^{\delta_i} \left(\underbrace{e^{-\lambda T_i^c}}\right)^{1-\delta_i} = \prod_{i=1}^n \lambda^{\delta_i} e^{-\lambda T_i^c}$$

log-vraisemblance :

$$\sum_{i=1}^n \delta_i \log \lambda - \lambda T_i^c = \log \lambda \left(\sum_{i=1}^n \delta_i\right) - \lambda \left(\sum T_i^c\right)$$

$\downarrow$ dérivation

$$\frac{\sum_{i=1}^n \delta_i}{\lambda} - \sum T_i^c \xrightarrow{EMV} \hat{\lambda} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n T_i^c}$$

Remarque : s'il n'y a pas de censure alors

$$\hat{\lambda} = \frac{\sum_{i=1}^n 1}{\sum T_i} = \frac{n}{\sum T_i}$$ cas de la stat. classique

$$\frac{1}{\hat{\lambda}} = \frac{\sum T_i^c}{\sum \delta_i}$$
censure

$$\boxed{\frac{1}{\hat{\lambda}} = \frac{\sum T_i}{\boxed{n}}}$$
sans censure.

$$\sum T_i^c \leq \sum T_i$$

$$\sum_{i=1}^n \delta_i \leq n$$

Slide 27.

$\boxed{(T^c, \delta)}$ on voudrait estimer $\bar{F}$, $\boxed{\lambda}$

survie hazardrate de la v.a. T,

$(\star\star) = \mathbb{P}(t \leq T^c \leq t+h, \delta = 1 \mid T^c \geq t)$

$$= \frac{\mathbb{P}(t \leq T^c \leq t+h, \delta=1)}{\mathbb{P}(T^c \geq t)} = \frac{\boxed{\mathbb{P}(t \leq T \leq t+h, T \leq C)}\ (\star)}{\boxed{\mathbb{P}(\min(T, C) \geq t)}}$$

$$\mathbb{P}(\min(T,C) \geq t) = \mathbb{P}(T \geq t, C \geq t)$$
$$= \mathbb{P}(T \geq t)\,\mathbb{P}(C \geq t)$$
$$= \boxed{\bar{F}(t)\,\bar{G}(t)}$$

$\boxed{T \perp\!\!\!\perp C}$

$T \sim f$

$C \sim g$.

$$(\star) = \frac{d}{dt}\mathbb{P}(T^c \leq t, \delta = 1) = f(t)\,\bar{G}(t)$$

$$(\star\star) = \boxed{\frac{f(t)\cancel{\bar{G}(t)}}{\bar{F}(t)\cancel{\bar{G}(t)}}} = \frac{f(t)}{\bar{F}(t)} = \boxed{\lambda(t)}.$$

$$\text{“}\overbrace{\mathbb{P}(t_i \leq T^c \leq t_i + h, \delta = 1)}\text{”} = \frac{\delta_i}{n}\text{”}$$

ti observé dans l'échantillon

$$\text{“}\overbrace{\mathbb{P}(T_i^c \geq t_i)} = \frac{n-(i-1)}{n}.$$

$$\text{“}\hat{\lambda}(t_i)\text{”} = \frac{\delta_i/n}{\frac{n-(i-1)}{n}} = \frac{\delta_i}{n-i+1}.\ \Bigg] \Leftarrow \text{“}\hat{f}(t_i) = \frac{1}{n}\text{”}$$

On utilise le résultat de la slide 11.

$$\hat{\bar{F}}(t_i) = \prod_{j=1}^{i}\left(1 - \hat{\lambda}(t_j)\right).$$