Application de la théorie des records sur la COVID-19 au Liban: prévision et prévention

Zaher KHRAIBANI

Lebanese University-Faculty of Sciences

December 10,2020

Université Libanaise

Objectif

- Détecter l'émergence d'une nouvelle maladie pour laquelle on a peu d'observations (cas d'une première émergence).
- Déterminer s'il s'agit d'une maladie sporadique ou d'une maladie émergente en se basant surl'approche par processus de records.

Qu'est ce qu'une maladie émergente ?

- Maladie dont l'incidence réelle augmente de manière significative dans une population, période donnée (début d'épidémie).
- Infections nouvelles, causées par l'évolution ou la modification d'un agent pathogène ou d'un parasite existant.
- L'événement de démarrage de l'émergence est la transition de la stabilité de l'état 0 pathogène à l'instabilité de cet état.

Définition mathématique d'une maladie émergente

- Soit Y_n le pourcentage de cas cliniques dans une population au temps n, tel que $Y_n = f(Y_{n-1})$ et $f(0) = 0$. En utilisant le développement de Taylor a l'ordre 1 au voisinage de 0 et en supposant Y_0 petit:

Définition mathématique d'une maladie émergente

- Soit Y_n le pourcentage de cas cliniques dans une population au temps n, tel que $Y_n = f(Y_{n-1})$ et $f(0) = 0$. En utilisant le développement de Taylor a l'ordre 1 au voisinage de 0 et en supposant Y_0 petit:
- $f(Y_0) = f(0) + Y_0 f^{'}(0) + O(Y_0^2) \Rightarrow f(Y_0) \approx Y_0 f^{'}(0)$.

Définition mathématique d'une maladie émergente

- Soit Y_n le pourcentage de cas cliniques dans une population au temps n, tel que $Y_n = f(Y_{n-1})$ et $f(0) = 0$. En utilisant le développement de Taylor a l'ordre 1 au voisinage de 0 et en supposant Y_0 petit:
- $f(Y_0) = f(0) + Y_0 f^{'}(0) + O(Y_0^2) \Rightarrow f(Y_0) \approx Y_0 f^{'}(0)$.
- $Y_1 = f(Y_0) = Y_0 f^{'}(0) + O(Y_0^2)$

Définition mathématique d'une maladie émergente

- Soit Y_n le pourcentage de cas cliniques dans une population au temps n, tel que $Y_n = f(Y_{n-1})$ et $f(0) = 0$. En utilisant le développement de Taylor a l'ordre 1 au voisinage de 0 et en supposant Y_0 petit:
- $f(Y_0) = f(0) + Y_0 f^{'}(0) + O(Y_0^2) \Rightarrow f(Y_0) \approx Y_0 f^{'}(0)$.
- $Y_1 = f(Y_0) = Y_0 f^{'}(0) + O(Y_0^2)$
- Donc $Y_2 = f(f(Y_0)) = f(Y_0)f^{'}(0) + O(f(Y_0)^2) = Y_0(f^{'}(0))^2 + f(Y_0)O(Y_0^2) + O(Y_0^2)$ Pour $f'(0) < 1$,

Définition mathématique d'une maladie émergente

- Soit Y_n le pourcentage de cas cliniques dans une population au temps n, tel que $Y_n = f(Y_{n-1})$ et $f(0) = 0$. En utilisant le développement de Taylor a l'ordre 1 au voisinage de 0 et en supposant Y_0 petit:
- $f(Y_0) = f(0) + Y_0 f^{'}(0) + O(Y_0^2) \Rightarrow f(Y_0) \approx Y_0 f^{'}(0)$.
- $Y_1 = f(Y_0) = Y_0 f^{'}(0) + O(Y_0^2)$
- Donc $Y_2 = f(f(Y_0)) = f(Y_0)f^{'}(0) + O(f(Y_0)^2) = Y_0(f^{'}(0))^2 + f(Y_0)O(Y_0^2) + O(Y_0^2)$ Pour $f'(0) < 1$,
- Dans le cas de non émergence:

$$f'(0) < 1 \Rightarrow f(Y_0) < Y_0 \Rightarrow Y_1 < Y_0$$

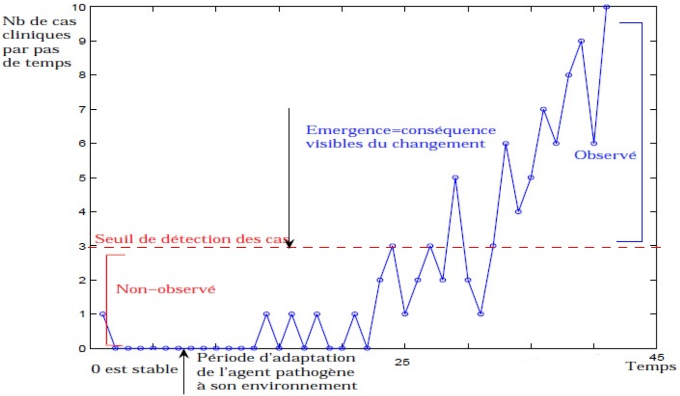
et $f(Y_1) = f(f(Y_0)) \approx f(Y_0)f^{'}(0) < Y_0$. On obtient par récurrence que $Y_n := f(n)(Y_0)$ reste du même ordre de grandeur que Y_0.

- Dans le cas d'une émergence: $f^{'}(0) > 1$ et on obtient :

$$f(Y_0) > Y_0 \Rightarrow Y_1 > Y_0$$

donc Y_n ne reste pas négligeable.

Figure 1: Émergence d'une maladie



Orientations méthodologiques

- Méthodes classiques pour détecter l'émergence ou la reémergence :

  - Méthodes asymptotiques (Colses,S.2001).

  - Méthodes descriptives (lissages, Tager IB 1996).

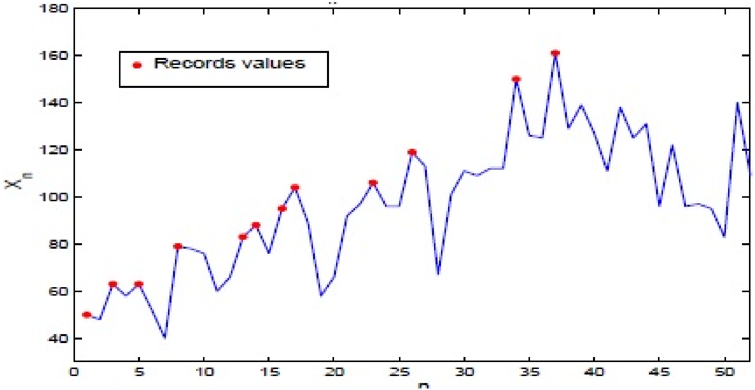
  - Méthodes exactes (tests non paramétriques (Run) William H. 2008).

Pourquoi la Théorie de Record?

- Les études théoriques des records a démarré en (1952, Chandler).
- Les records jouent un rôle important dans plusieurs domaines comme le changement climatique, les catastrophes naturelles, Sports,Economie et Finance...
- Robustesse et lois exactes.
- Information sur les instants des records $\{L_n\}_{n\geq 1}$, non seulement a l'étude de la loi limite du maximum $\{M_n = \max(X_1; ...; X_n)\}_{n\geq 1}$.

Représentation des valeurs de records

- Le processus de records représente la tendance maximale observée de l'épidémie.



Définition du processus de records

- Par Définition, $\{R_n : n \geq 1\}$ et $\{L_n : n \geq 1\}$ sont respectivement la suite des valeurs des records et la suite des indices des records. Plus précisément :

$$
\begin{aligned}
L_1 &= 1 \\
L_n &= \inf\{j > L_{n-1} : X_j > X_{L_{n-1}}\} \\
R_n &= X_{L_n}
\end{aligned}
$$

- Soit N_n le nombre total des records parmi $\{X_1; ...; X_n\}$ avec $N_1 = 1$:

$$N_n = \sum_{j=1}^{n} \delta_j;$$

oú δ_j est l'indicatrice de record tel que :

$$\delta_j = \begin{cases} 1, & \text{si } X_j > max(X_1, ..., X_{j-1}) \\ 0, & \text{sinon} \end{cases}$$

Nombre de records

Les principaux résultats de la Théorie de Record dans le cas i.i.d. ont été obtenus au cours de la période 1952-1983 (Voir Chandler (1953), Arnold (1998) et Nevzorov (2001)):

- Les v.a. $\{\delta_n\}_{n\geq 1}$ sont indépendantes et $\delta_n \sim \textit{Bernoulli}(1/n)\textit{avec}$:

$$\begin{aligned} P_n &= \text{taux de record (la probabilité que } X_n \text{ soit un record)} \\ &= \mathbb{P}[\delta_n = 1] \\ &= 1/n \end{aligned}$$

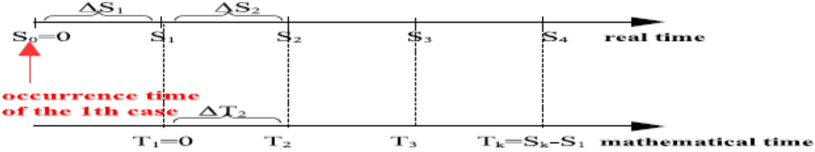
- La distribution exacte de N_n est donnée par:

$$\mathbb{P}[N_n = m] = \frac{s(n,m)}{n!},\ 0 \leq m \leq n$$

- En plus Arnold et Nevzorov donnent les distributions marginales et jointes des suites $\{R_n\}_{n\geq 1}$ et $\{L_n\}_{n\geq 1}$.

Formalisation mathématique du problème

Figure 2: Processus Ponctuel



- $\{S_n\}_{n\geq 0}$: instants d'occurrence successifs des cas cliniques (processus de renouvellement).
- $X_n = (\Delta S_n)^{-1}$, $\Delta S_n = S_n - S_{n-1}$,temps d'attente entre l'arrivée de deux cas successifs. Dans le cas d'une émergence $\{\Delta S_n\} \searrow \Leftrightarrow \{(\Delta S_n)^{-1}\} \nearrow \Rightarrow$ records de $X_n \Rightarrow \{\mathrm{T}_n, X_n\}$: Processus Ponctuel.

Formalisation des hypothèses à tester

1. H_0 (cas sporadiques): $\{\Delta S_n\}$ i.i.d,

   $P(\Delta S_n \leq s) := E(s) = 1 - \exp(-\lambda s)$ (loi sans mémoire).

Formalisation des hypothèses à tester

1. H_0 (cas sporadiques): $\{\Delta S_n\}$ i.i.d,

   $P(\Delta S_n \leq s) := E(s) = 1 - \exp(-\lambda s)$ (loi sans mémoire).

2. H_1 (émergence): $\{\Delta S_n\}$ indépendantes, $\bar{E}_n = \bar{E}_n{}^{\rho_n}$,

   $(\bar{E}_n = 1 - E)$, $\{\rho_n\}_n$ suite positive croissante.

Loi du nombre de records sous H_0 et H_1

Proposition

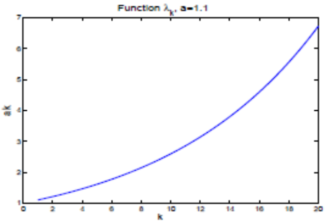
$P_{H_0 \cup H_1}(N_n = m) = \frac{|s(n+1,m+1|\vec{u})|}{\prod_{j=1}^{n+1}(1+u_{j-1})}$ où $s(n+1, m+1|\vec{u})$ (nombre de Stirling généralisé de première espèce), $\vec{u} = (u_0, ..., u_n)$, $u_{j-1} = \frac{\sum_{k=1}^{j-1} \rho_k}{\rho_j}$, $j \geq 1$.

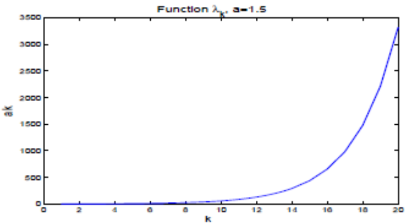
- Cas particulier: $|s(n+1, m+1|\vec{u})| = s(n+1, m+1)$, si $\rho_k = \rho, \forall k$,
- $P_{H_0}(N_n = m) = \frac{s(n+1,m+1)}{(n+1)!}, 0 \leq\leq n$, avec $\{s(n+1, m+1)\}$ le nombre de Stirling de première espèce.

Fréquence de l'émergence

- $E(\Delta T_k) = (\lambda_k)^{-1}$ où $\lambda_k = \lambda.\rho_k = \lambda.a^k$ est la fréquence du nombre de cas par unité de temps au temps T_k ,$a > 1$, croissance exponentielle d'une maladie infectieuse.

Figure 3: Fonction $\{\lambda_k\}$ où $\lambda_k = a^k$, pour $a = (1.1, 1.5)$ et $\lambda = 1$.





Application du test de records sur le COVID-19

- Test de H_0 contre H_1: définitions et Rappels

  Les hypothèses à tester:

  -H_0 (maladie sporadique): $\{X_k\}$ i.i.d., $X_1 \sim F$, F continue.

  -H_1 (maladie émergente) : $\{X_k\}$ indépendantes $X_k \sim F_k$ où

  $\bar{F_k} = \bar{F}^{\rho_k}$.

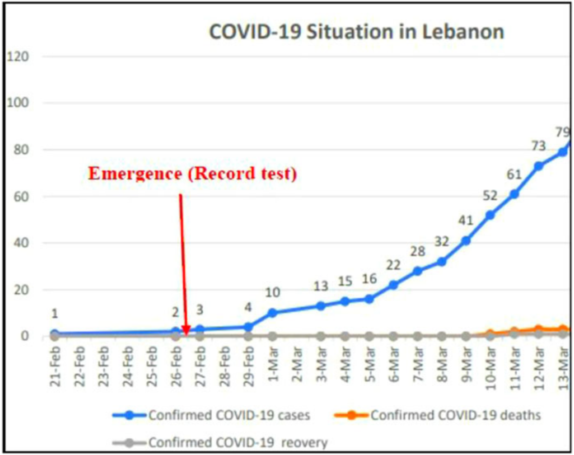
- Statistique de test utilisée : N_n.

- Risque d'erreur: $\alpha = P_{H_0}(rejeter\ H_0) = P_{H_0}(N_n \geq N_\alpha)$, (la

  région de rejet de H_0 est déterminée par les grandes valeurs de

Illustration graphique

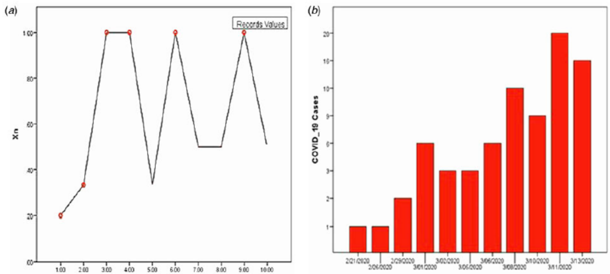
Figure 4: Daily cumulative/emerged number of confirmed, fatal and recovered cases of Coronavirus Disease 2019 (COVID-19) in Lebanon.



Description

- Nouvelle maladie, observée dans différents pays, pas d'information épidémiologique.
- 10 cas détectés au Liban de 21 Fev.2020 au 1 Mars 2020.
- ΔT_n, le temps d'interarrivée entre le *nième* et le $(n+1)$*ième* cas.

Figure 5: (a) Records values of $X_n = (\Delta T_n)^{-1}$. (b) Number of observed COVID-19 cases per day.



Valeurs et nombre de records

Table 1. Waiting times between two successive cases and number of COVID-19 cases in Lebanon per day

T_n	21/02	26/02	29/02	01/03	02/03	05/03	06/03	08/03	10/03	11/03	13/03
COVID	1	1	2	6	3	3	6	10	9	20	16
(ΔT_n)	5	3	1	1	3	1	2	2	1	2	–
$(\Delta T_n)^{-1}$	0.2	0.33	1	1	0.33	1	0.5	0.5	1	0.5	–

- Le nombre de records observés $N_n = 6$

Table 2. $P(N_n \geq m)$, for $n = 10$, 20 and for different values of m

	m						
n	1	2	3	4	5	6	7
10	1	0.9000	0.6171	0.2939	0.0945	0.0203	0.0029
20	1	0.9500	0.7726	0.4978	0.2470	0.0944	0.0280

- $P_{H_0}(N_{10} \geq 6) = 0.0203 \Rightarrow$ Rejet H_0, (À partir de quelques cas observes au début de l'épidémie, nous concluons que le COVID-19 est une maladie émergente au Liban).

Test H_0, H_1

Table 3. $P(N_n \geq m)$ under H_1, for $n = 10$, 20, 30 and for different values of m and a

N	10	20	30
m	6	7	8
$P(N_n \geq m)$, $a = 1.1$	0.0455	0.1108	0.14261
$P(N_n \geq m)$, $a = 1.5$	0.2683	0.7554	0.9352

- On pourrait remarquer que le COVID-19 au Liban émerge très rapidement avec une forte probabilité de dépasser certaines valeurs records pour $n = 10, 20, 30$.

Prévision de l'émergence

- En prenant en compte les futurs COVID-19 records, nous calculons la probabilité de temps d'attente ($\Delta T_n{}^*$) pour observer un nouveau records:

$$P(\Delta T_n{}^* > n^*) = \frac{n^{'}}{n^{'} + n^*}$$

- Pour ($n^* = 5$) et $n^{'} = 22$ jours (21 février 2020-13 mars 2020):
- $P(\Delta_n{}^* > 5) = 22/(22+5) \approx 0,82 \Rightarrow$ Croissance rapide de la fréquence du COVID-19 au Liban sur une courte période.

Conclusion

1. Nouvelle méthodologie pour détecter l'émergence d'une nouvelle maladie.
2. Tester H_0 (maladie sporadique) contre H_1 (émergence de la maladie).
3. N_n: statistique de test (robuste) indépendante de F sous H_0 et H_1, exactement calculée pour chaque valeur de n.
4. Sous H_1 la distribution du N_n dépend seulement du paramètre $\rho_k = a^k$ (croissance exponentielle du nombre de cas d'une maladie infectieuse).
5. L'approche par les records est particulièrement adéquate pour n petit.

Références

1. Ahsanullah M (2004) Record Values-Theory and Applications. Oxford: University Press of America.
2. Khraibani Z et al.(2015). A non parametric exact test based on the number of records for an early detection of emerging events: illustration in epidemiology. Communications in Statistics-Theory and Methods 44, 726–749.
3. Khraibani, Z. et al.(2020). Application of records theory on the COVID-19 pandemic in Lebanon: Prediction and prevention. Epidemiology and Infection, 148, E192. doi:10.1017/S0950268820001909.
4. Kucharski A et al.(2020). Early dynamics of transmission and control of 2019-nCoV: a mathematical modelling study. Lancet Infectious Diseases 20, 553–558.
5. Official website of Johns Hopkins University.Available at https://gisanddata.maps.arcgis.com/apps/.
6. World Health Organization. WHO characterizes COVID-19 as