

R

ENDEZ-VOUS

P.80 Logique & calcul
P.86 Art & science
P.88 Idées de physique
P.92 Chroniques de l'évolution
P.96 Science & gastronomie
P.98 À picorer

PAS SI SIMPLE, L'ESPÉRANCE DE VIE!

L'espérance de vie publiée par les statisticiens indique-t-elle combien de temps en moyenne il nous reste à vivre? Non, car les calculs reposent sur les statistiques passées, qui ne se reproduiront pas nécessairement...

L'AUTEUR



JEAN-PAUL DELAHAYE
professeur émérite
à l'université de Lille
et chercheur au
laboratoire Cristal
(Centre de recherche
en informatique, signal
et automatique de Lille)



Jean-Paul Delahaye
a notamment publié :
**Les Mathématiciens
se plient au jeu,**
une sélection de ses
chroniques parues
dans *Pour la Science*
(Belin, 2017).

L'espérance de vie est un concept statistique intéressant... et mal compris. Il n'est pas vrai par exemple que si l'espérance de vie à la naissance en France publiée par l'Insee dans ses tableaux 2020 de l'économie française (voir la bibliographie) est de 85,6 ans pour les femmes, cela signifie que les femmes nées en France en 2020 vivront en moyenne 85,6 ans. Le sens des données publiées est plus subtil!

L'ESPÉRANCE MATHÉMATIQUE

Voyons sur un exemple ce que signifie l'«espérance mathématique». Imaginons que vous lancez un dé et que cela vous rapporte en euros le résultat du lancer: 1 euro si le 1 sort, 2 euros si le 2 sort, etc. Combien gagnez-vous par lancer en moyenne?

C'est facile: une fois sur six vous gagnerez 1 euro, une fois sur six, 2 euros, etc. Comme les chances de chaque résultat sont égales, vous gagnerez la moyenne de leurs valeurs, soit $(1+2+3+4+5+6)/6 = 21/6 = 3,5$ euros, calcul que l'on peut voir comme la moyenne des sommes gagnées pondérées par leurs probabilités: $(1 \times 1/6) + (2 \times 1/6) + (3 \times 1/6) + (4 \times 1/6) + (5 \times 1/6) + (6 \times 1/6) = 21/6 = 3,5$ euros. Ce sera votre espérance de gain à chaque lancer de dé.

Imaginons maintenant que le dé est truqué: le 6 tombe une fois sur deux, et le 1, le 2, le 3, le 4 et le 5 tombent chacun une fois sur dix. Le total des probabilités est bien égal à 1:

$1/10 + 1/10 + 1/10 + 1/10 + 1/10 + 1/2 = 1$, tout va bien. Votre gain moyen au jeu envisagé plus haut est toujours la moyenne des sommes gagnées en les pondérant par leurs probabilités, ce qui donne: $(1 \times 1/10) + (2 \times 1/10) + (3 \times 1/10) + (4 \times 1/10) + (5 \times 1/10) + (6 \times 1/2) = 4,5$ euros. Le dé truqué vous fera gagner en moyenne 1 euro de plus par lancer que le dé équilibré.

Ainsi, l'espérance pour une variable aléatoire (ici le résultat du lancer d'un dé) est la valeur moyenne obtenue en multipliant les valeurs possibles par les probabilités d'avoir ces valeurs. L'espérance de vie à la naissance pour une catégorie de personnes, par exemple celles nées en 2001, est la moyenne des âges de décès de ces personnes, pondérés par leurs probabilités respectives, c'est-à-dire par la fréquence mesurée de chacun de ces âges de décès. C'est l'«espérance de vie réelle» (EV_R). Or ce n'est pas cela que les statisticiens appellent l'«espérance de vie»!

INACCESSIBLE ESPÉRANCE DE VIE RÉELLE

On ignore en effet à quel âge mourront les gens qui naissent cette année. L'espérance de vie réelle à la naissance pour la catégorie de personnes nées durant l'année N n'est calculable qu'un siècle environ après l'année N . Sauf pour les générations très anciennes, par exemple les Français nés en 1910, elle ne peut

pas être déterminée. On aimerait quand même avoir une idée de cette espérance de vie, y compris pour les gens nés en 2021 ou en 1990! Comment tirer parti des statistiques passées?

Une première idée serait de calculer l'âge moyen de décès (AMD), durant l'année N , des personnes auxquelles on s'intéresse, par exemple les Françaises durant l'année 2021. Si la structure des âges et les taux de mortalité dans la population étudiée étaient constants dans le temps, cette méthode serait satisfaisante. Ce n'est pas le cas.

L'idée du calcul de l'Insee et des statisticiens dans tous les pays est plus subtile: ils calculent les taux de mortalité par âge à partir des données les plus récentes disponibles (celle de 2019 par exemple), puis ils imaginent une population fictive qui évolue d'année en année en subissant ce taux de mortalité pour chaque âge. Les enfants de moins d'un an de cette population fictive meurent en proportion comme l'indique le taux de mortalité pour 2019 des enfants de moins d'un an; puis les enfants ayant survécu à leur première année meurent dans leur seconde année avec le taux de mortalité pour 2019 des enfants âgés entre un et deux ans, etc. L'espérance de vie mathématique pour cette population fictive est alors facile à calculer: c'est la moyenne des âges de décès pondérés par les effectifs associés. Le résultat de ce calcul est ce que les statisticiens publient, et nous le nommerons «espérance de vie des statisticiens» (EV_s).

Résumons: l'espérance de vie réelle (EV_r) fondée sur la notion mathématique d'espérance (comme pour les dés) n'est calculable que pour les générations anciennes, aussi on la remplace par le calcul de EV_s , qui se fonde chaque année sur les chiffres des statistiques de décès de la dernière année pour laquelle on en dispose, en imaginant une génération fictive respectant ces taux de mortalité. Les encadrés 1 et 2 donnent des exemples montrant numériquement les différences entre EV_r et EV_s . Cela étant, quels risques prend-on en calculant EV_s à la place de EV_r ?

POPULATIONS IMAGINAIRES, MAIS POSSIBLES

L'espérance de vie à la naissance n'est pas l'information qui vous concerne le plus, c'est plutôt «l'espérance de vie à l'âge A », quand vous avez l'âge A .

Dans le cas de EV_r , ce sera l'espérance au sens mathématique du nombre d'années de vie restant à vivre pour une personne d'âge A , en se référant aux chiffres réels des décès de gens ayant l'âge A à la date à laquelle on s'intéresse (par exemple $N = 1880$), qui ne sont connus que progressivement quand les années passent.



ESPÉRANCE DE VIE ET ÂGE MOYEN DE DÉCÈS

L'espérance de vie réelle (EV_r), l'espérance de vie des statisticiens (EV_s) et l'âge moyen de décès (AMD) sont différents. Imaginons un élevage animal dont les effectifs par âge aux 1^{er} janvier 1900, 1901, 1902 et 1903 sont donnés par le tableau :

Âge (ans)	0	1	2	3	4
Effectif 1900	100	100	100	100	0
Effectif 1901	200	80	60	50	0
Effectif 1902	300	170	56	30	0
Effectif 1903	400	270	136	28	0

On en déduit les taux de décès par âge (TCA) chaque année :

Âge (ans)	0	1	2	3
TCA en 1900	20%	40%	50%	100%
TCA en 1901	15%	30%	50%	100%
TCA en 1902	10%	20%	50%	100%

Nous avons supposé que le nombre de naissances augmentait globalement, ce qui se produit en général pour les populations humaines. Ici, pour un élevage, ce nombre passe de 100 à 400. Nous avons aussi supposé que les taux de mortalité étaient chaque année croissants en fonction de l'âge, mais que, d'une année à la suivante, ils décroissaient avec les progrès dans les soins apportés aux animaux de l'élevage. Pour les humains, il y a aussi une augmentation des taux de mortalité avec l'âge et une baisse de ces taux d'année en année avec les progrès de la médecine et de l'hygiène.

Espérance de vie EV_r à la naissance pour les animaux d'âge 0 en 1900

Parmi les 100 individus nés en 1900, 20 sont morts avant d'atteindre l'âge de 1 an, 24 sont morts dans leur seconde année, c'est-à-dire quand ils avaient 1 an, 28 sont morts quand ils avaient 2 ans, et 28 quand ils avaient 3 ans. L'espérance de vie pour cette cohorte née en 1900 est donc : $EV_r = 0 \times 20/100 + 1 \times 24/100 + 2 \times 28/100 + 3 \times 28/100 = 1,64$ an. Pour la connaître, il faut attendre la

mort de tous les animaux concernés, mais c'est la véritable espérance de vie : la moyenne des âges de décès pondérés par les probabilités de chacun de ces âges de décès.

Espérance de vie EV_s à la naissance pour les animaux d'âge 0 en 1900 évaluée à partir des taux de mortalité de 1900

Si nous calculions l'espérance de vie des animaux de l'élevage à la fin de l'année 1900, nous devrions utiliser les taux de mortalité pour l'année 1900. C'est la situation des statisticiens qui, en 2021, veulent calculer l'espérance de vie des enfants qui viennent de naître en ne disposant que des données de 2020 ou même de 2019. Pour l'élevage, le statisticien fait l'hypothèse que les taux de mortalité de 1900 seront valides tout au long de la vie des animaux nés en 1900.

Les taux de mortalité en fonction de l'âge en 1900 sont 20 %, 40 %, 50 %, 100 %. La population fictive imaginée par le statisticien pour le calcul de l'espérance de vie passera donc de 100 individus à 80 entre 1900 et 1901, puis de 80 à 48 entre 1901 et 1902, puis de 48 à 24 entre 1902 et 1903, puis disparaîtra. Cela donne : $EV_s = 0 \times 20/100 + 1 \times 32/100 + 2 \times 24/100 + 3 \times 24/100 = 1,52$ an. L'espérance de vie du statisticien (EV_s) est inférieure à la réalité (EV_r), car le statisticien, avec les données de 1900, a surestimé les taux de mortalité, qui, grâce au progrès de l'élevage, ont diminué d'année en année.

Âge moyen de décès en 1900

Il y a eu 210 morts en 1900, et l'âge moyen des animaux décédés (AMD) est : $AMD = 0 \times 20/210 + 1 \times 40/210 + 2 \times 50/210 + 3 \times 100/210 = 2,09$ ans. Cet âge est encore plus élevé que EV_r , car les données pour 1900 contiennent trop d'animaux âgés, ce qui se corrige pour 1901, 1902 et 1903. Un plus grand nombre de sujets âgés augmente l'âge moyen de décès, comme un plus grand nombre de jeunes le ferait décroître : l'âge moyen de décès ne donne pas l'espérance de vie...