*Cours d'analyse démographique* (**Master 1e année**) par Alexandre Avdeev

Chapitre 2

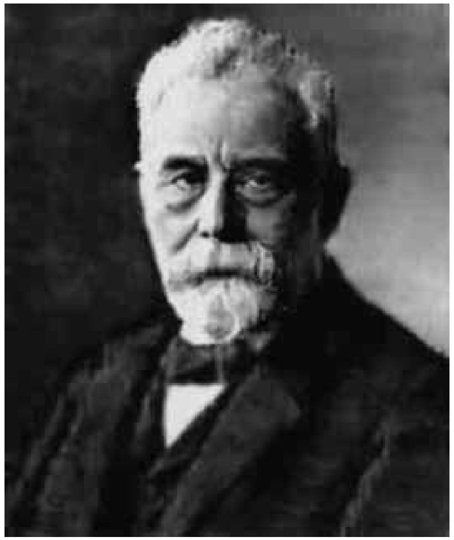
Diagramme de Lexis, types des taux et des quotients par âge

- Diagramme de W. Lexis : version d'origine et les modifications récentes
    - localisation des populations soumises au risque
    - localisation des événements démographique dans le temps
- Notion de temps dans l'histoire de vie
- Représentation graphique des parcours et des événements
- Type des taux et calculs des taux,
- Données nécessaires et leur localisation sur le diagramme de Lexis,
- Comparaison et relations entre les taux de divers types,
- Quotients, ou les probabilités d'une événement,
- Taux/quotients de première et de deuxième catégorie.

Lecture : Grazielle Cazelli et Jacques Vallin « Variation dans le temps des taux par âge » // *Démographie : analyse et synthèse*, vol. I La dynamique des populations. (sous la dir. De G.Cazelli, J.Vallin et G.Wunsch), Paris, INED, 2001, p.85-111

Méthode graphique de repérer les événements démographiques et les populations correspondants dans le temps

Le diagramme de Wilhelm Lexis (1837-1914)



À propos de quelques particularités des taux et quotients

- Le quotient est une proportion de personnes ayant eu l'expérience d'un événement dans la population exposée au risque *au début d'une période* déterminée : *plus la période est longue, plus la probabilité de l'événement est grande* **(s'il est question d'un événement fatal, i.e. décès, alors avec t→ ∞, q → 1).**
- Le quotient *est censé de mesurer la probabilité* dans le sens **de sa définition classique** (quand le nombre possible d'événements d'intérêt est strictement défini), or la probabilité a des propriétés (les probabilités indépendantes sont additives, les probabilités conditionnelles sont multiplicatives etc.), on attribue donc les propriétés de la probabilité aux quotients.
- En réalité, quotient n'est qu'une intégrale (*la primitive*) de la fonction de densité de la probabilité (dont le taux est une mesure de la tendance centrale avec une hypothèse sous-jacente que sa distribution est linéaire ou uniforme sur l'intervalle d'intégration). Par conséquent, le score d'un quotient dépend de l'amplitude de l'intervalle d'intégration. Donc ce n'est pas un indicateur robuste (en plus la vraie distribution de la densité est inconnue dans la plupart des cas) .
- Pour calculer un quotient on suppose :
  - ✓ que le temps soit discret (et en réalité ce n'est jamais le cas) ;
  - ✓ qu'une population soit homogène (une condition qui est très difficile à remplir)

*p.ex. l'indice de croissance de la population dans le temps discret représente de fait un quotient, mais cet indice est certainement influencé par **l'amortissement démographique** (changements dans les structures sous l'influence des événements démographiques) et probablement – par **des compensations démographiques** (accumulations sous influences non démographiques) dans une population hétérogène (voir les pyramides démographiques irrégulières comme celle de la Russie en 1959)*

- Pour remplir les conditions de calcul des quotients et pour atténuer les problèmes de la hétérogénéité:
  - ✓ on choisit des ***populations (groupes) très homogènes*** (un groupe d'âge → quotients de mortalité; célibataires à un âge donné → quotient de primo-nuptialité, etc.)
  - ✓ on fait accourcir le période (*t*) qui rarement dépasse une année;
    si *t* →*0*, quotient devient ≡ à un taux instantané

Taux/quotient de « seconde » catégorie : une simplification pour pallier le problème d’hétérogénéité

On considère comme les taux de « première » catégorie ceux qui rapportent le nombre d’événements démographiques ***E*** à une population réellement soumise au risque de cet événement P_e.
P.ex.: les mariages et les célibataires, les premières naissances et les femmes sans enfants etc.

Les calculs de ces taux ne sont pas toujours faciles à cause des autre événements qui entrent en concurrence avec les événements d’intérêt en les empêchant (p.ex. mariage et décès) ou en les perturbant (p.ex. mariage et migration).

On calculs les taux de « seconde » catégorie en rapportant le nombre d’événements démographiques l’effectif de la population soumise au risque de cet événement et non. Plus souvent à la population totale.

Les taux de « seconde » catégorie peuvent déformer fortement une image réelle des processus démographiques, puisque, par ex., ils sont chargés du passé (la natalité ou le mariage), mais ils sont souvent additifs, ce qui simplifie leur intégration dans le modèle additifs comme celui du bilan démographique.

$$\text{Taux} = \frac{\text{Nombre d'occurences durant une période T}}{\text{Nombre d'années vecues par la population exposée au risque et non}} = \frac{E}{P}$$

De fait, les taux de seconde catégorie sont plutôt les ratios

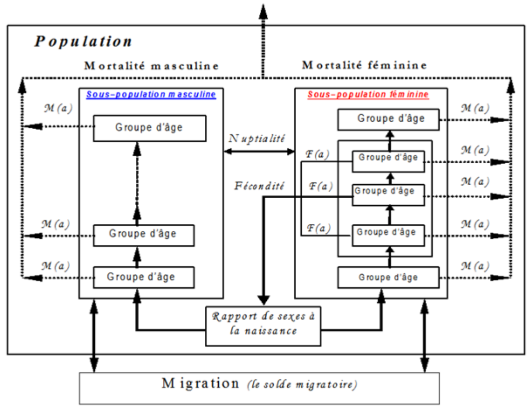
$$T_1 = \frac{E}{P_e} \qquad T_2 = \frac{E}{P}$$

Soit T_1 taux de première catégorie, T_2 taux de seconde catégorie
- ***E*** – nombre d’événements,
- P_e – population réellement exposée au risque d’événement ***e***,
- P_{-e} – population non exposée au risque ***e***,
- ***P*** – population totale exposée et non au risque ***e*** → $P=P_e+P_{-e}$

$$T_2 = T_1 \cdot \frac{P_e}{P} = \frac{E}{P_e} \cdot \frac{P_e}{P}$$

Le principe de correspondance ou la « règle d’or » de calcul des indicateurs démographiques: toujours rapporter les évènements à la population qui les a effectivement produits

- On a étudié les indicateurs du *mouvement général* de la population dont la construction est basée sur un concept de *la variation dans* ***le temps*** (croissance) et des composants de cette variation (bilan démographique)
- Ces indicateurs sont considérés comme étant « bruts » puisqu’ils sont calculés comme si la *population était un ensemble homogène* (réduit à un nombre) par rapport à l’événement d’intérêt, tandis qu’en réalité ce n’est pas vrai (***la moralité et la fécondité varient avec l’âge*** etc…)
- Pour calculer les indicateurs du mouvement d’une population, *il faut assumer que les sous-populations* (pour lesquelles on calcule ces indicateurs) *soient homogènes* relativement aux caractéristiques des événements qui sont caractérisés par ces indicateurs.
- On peut y rencontrer des problèmes puisque les données sur les ensembles d’évènements se réfèrent toujours à une période, tandis que les caractéristiques des populations se réfèrent à un moment.



Par ex. : taux de mortalité à l’âge x durant une période t :

$$m(x;t)=\frac{D(x;t)}{\overline{P}(x)\cdot t} \quad \text{avec} \quad \overline{P}(x)=\frac{P(x;t_0)+P(x;t_t)}{2}$$

Formellement, ici la condition de la homogénéité est respectée, quoi que P(x) soit un mélange de différentes générations ; et $\boldsymbol{P(x;t_0)}$ – population en début de période et $\boldsymbol{P(x;t_t)}$ – celle à la fin de période soient liées par l’équation du bilan démographique

Taux par âge

- On a vu que la structure par âge et par sexe était très variable (pyramidale)
- Notre expérience nous dit que la mortalité (probabilité de décéder) et la fécondité (probabilité d'accoucher un enfant) dépendant du sexe et varient avec l'âge

<u>Les taux calculés pour les intervalles d'âge :</u>

Soit $\boldsymbol{D^x_{0,t}}$ le nombre de décès à l'intervalle d'âge « ***x*** » durant une période entre ***0*** et ***t***;

$\boldsymbol{B^x_{0,t}}$ le nombre de naissance à l'âge « ***x*** » durant une période entre ***0*** et ***t***.

taux par âge

événement par âge

$$m^x_{0,t} = \frac{D^x_{0,t}}{t \cdot \bar{P}^x_{0,t}}$$

taux de mortalité pour âge « x » et période [0,t)

$$D^x_{0,t} = m^x_{0,t} \cdot t \cdot \bar{P}^x_{0,t}$$

$$f^x_{0,t} = \frac{B^x_{0,t}}{t \cdot \bar{P}^x_{0,t}}$$

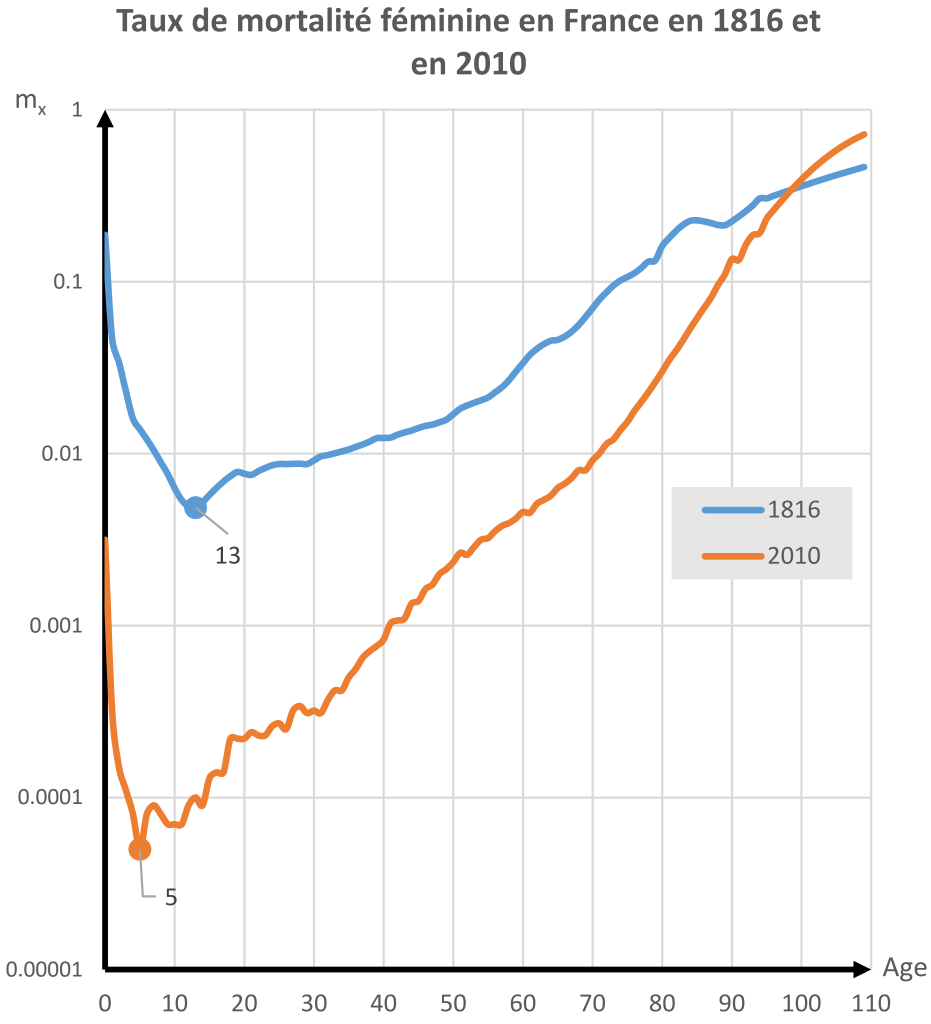
taux de fécondité pour âge « x » et période [0,t)

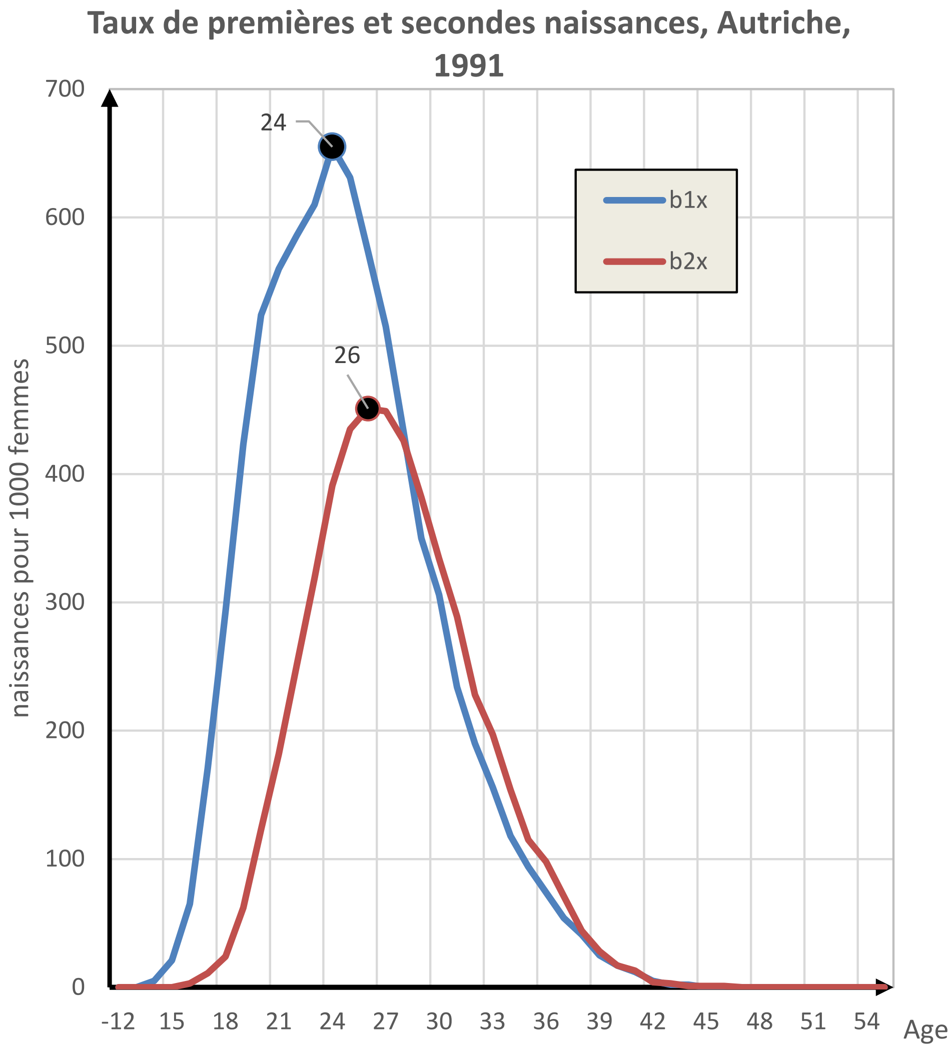
$$B^x_{0,t} = f^x_{0,t} \cdot t \cdot \bar{P}^x_{0,t}$$

Les sommes d'événements et, par conséquent, les taux bruts, dépendent des taux par âge

Etant donné que la durée de la période est une année (t = 1), nous pouvons nous passer du symbole *t* des formules

Taux ou quotient par âge (par groups d'âge)





Note: il est d'usage et autorisé par la tradition de présenter les répartitions des quotients sur le graphique en lignes continues, bien que du point de vue mathématique cela ne soit pas tout à fait convenable. Les quotients doivent plutôt être présentés comme un graphique en « marche d'escalier », ou comme un histogramme

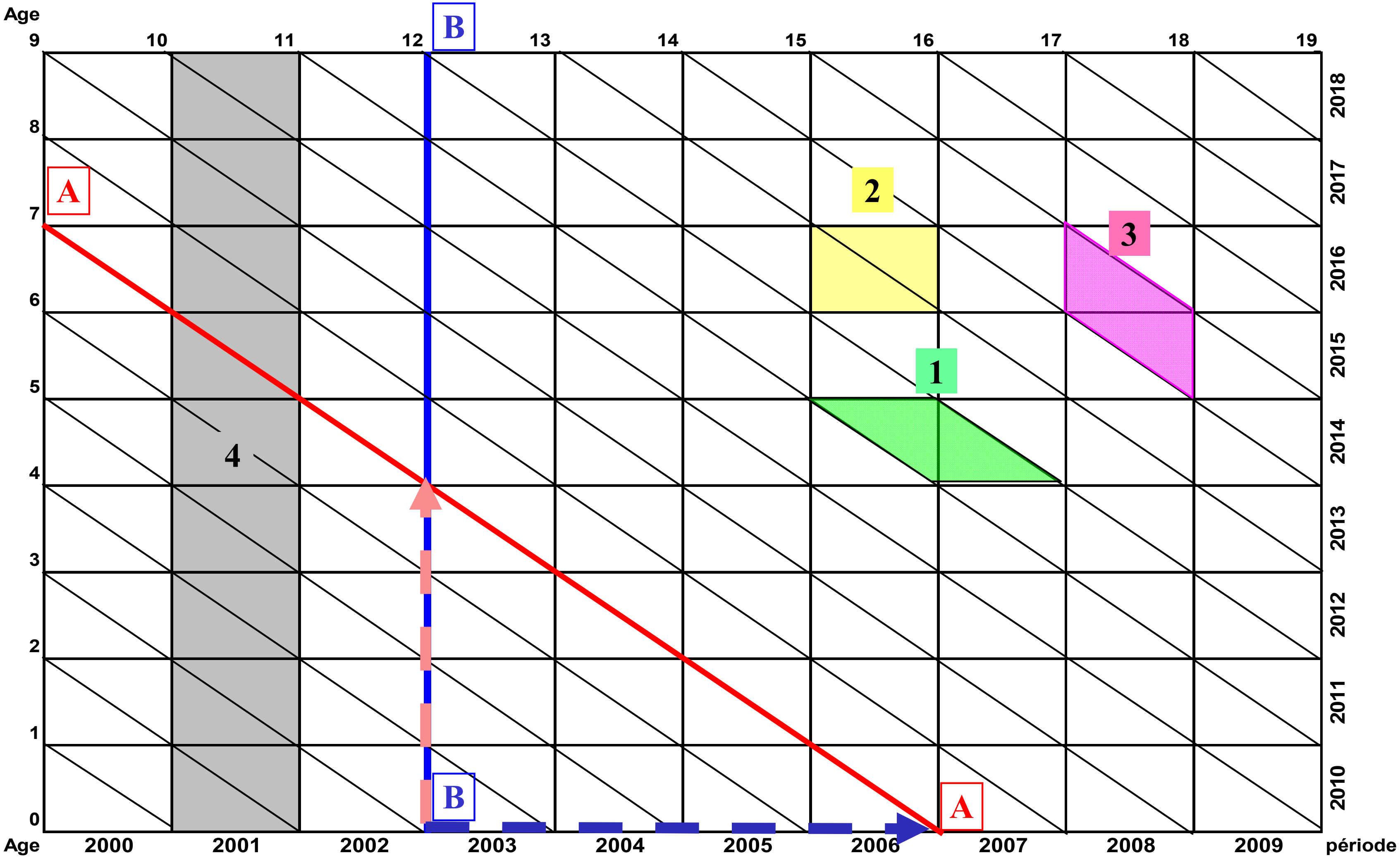
Notion de temps en analyse démographique: temps général et temps individuel

- Dès qu'on passe d'une <u>population globale, qui existe dans le temps, mais qui n'a pas d'âge</u> à des <u>sous-populations qui non seulement évoluent dans le temps, mais qui changent constamment leur âge</u>, nous somme obligé à préciser les notions de temps et de l'âge pour pouvoir mieux identifier des données démographiques.
- Mesure de temps = l'échelle d'intervalles (un enfant à l'âge de 10 ans a 5 ans de plus et deux fois plus âgé qu'un enfant à l'âge de 5 ans, et dans 50 ans ?)
- Hormis *un temps « physique » ou « astronomique », général pour tous*, nous devons définir un *temps « individuel » qui est l'âge*, ou le temps écoulé depuis un événement constituant (e.g. la naissance) et un moment de temps de calendrier choisi pour l'intérêt d'étude (e.g. fin de l'année civile).
- *Les deux temps sont continus*, mais on pourrait les quantifier convenablement par catégories d'unité de temps, (année d'âge et année civile), pour associer les indicateurs relatifs aux mouvements de la population aux intervalles de temps, le plus souvent les « annualiser » .
- *Les deux temps sont **isomorphes*** puisque un événement constituant est défini dans le temps de calendrier, <u>mais leurs scores annualisés ne coïncident qu'exceptionnellement</u> (on vive, le plus souvent, deux années d'âge durant une année civile et une année d'âge s'étend normalement sur deux années civiles, sauf si la naissance a eu lieu juste au début l'année civile)

Les ensembles démographiques sur le diagramme de Lexis (en VO)

l’axe horizontale = année de naissance; l’axe verticale = âge

Lexis W. (1875). *Einleitung in die Theorie der Bevölkerungsstatistik*. Straßburg: Karl J. Trübner.



A - isochrone (les événements qui se sont produisent à la même date)

B - ligne de vie d'une génération

1 - les événements: âge X période (sur deux générations consécutives)

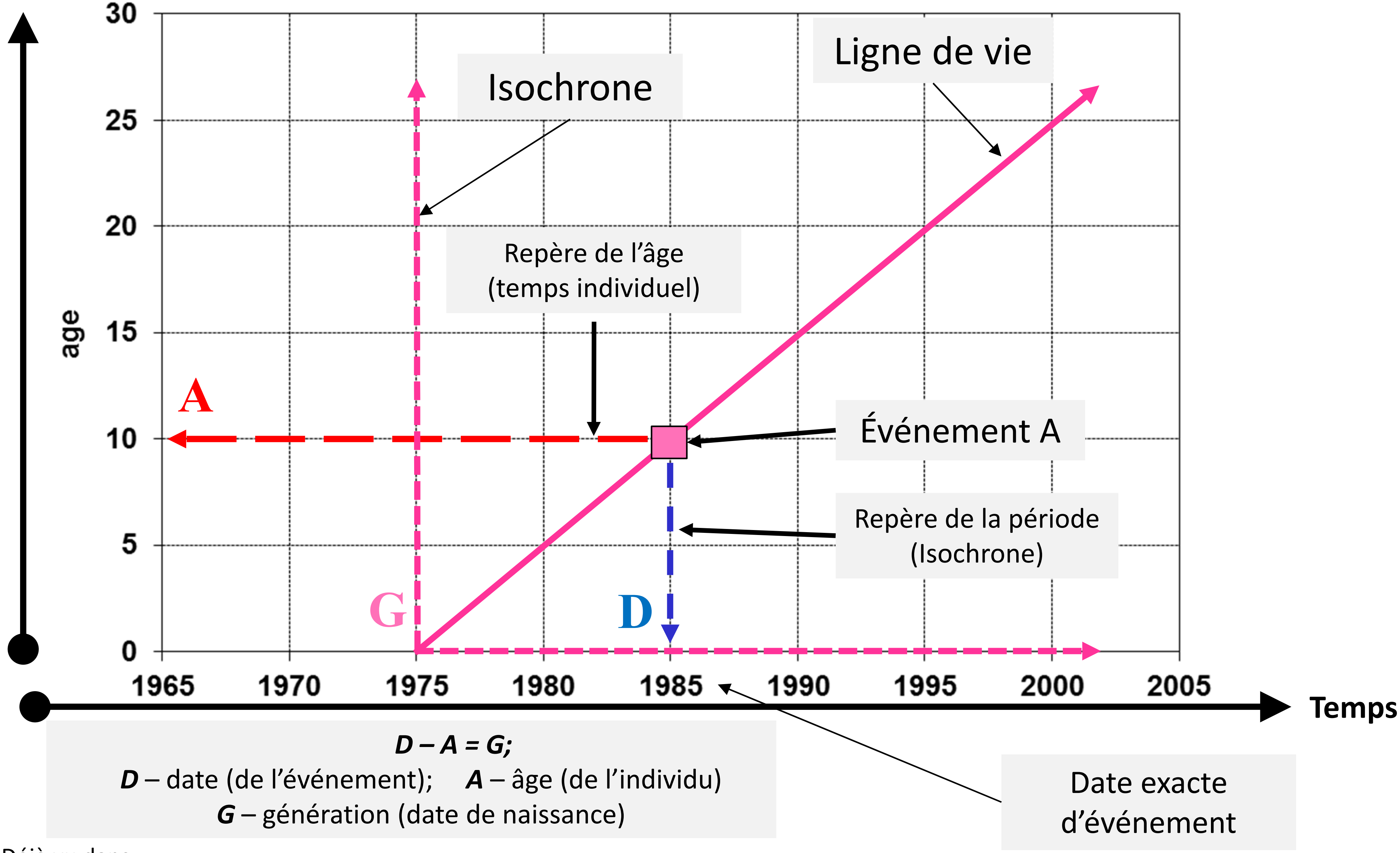
2 - les événements: génération X âge (sur deux années consécutives)

3 - les événements: génération X période (sur deux âges consécutifs)

4 - génération née entre le 1 janvier 2001 et le 1 janvier 2002 (au cours de l'an 2001)

Le diagramme Lexis-Pressat (modification de R. Pressat):

les isochrones et les lignes de vie sont inversées par rapport à la VO du diagramme



***D* – *A* = *G*;**

D – date (de l'événement); ***A*** – âge (de l'individu)

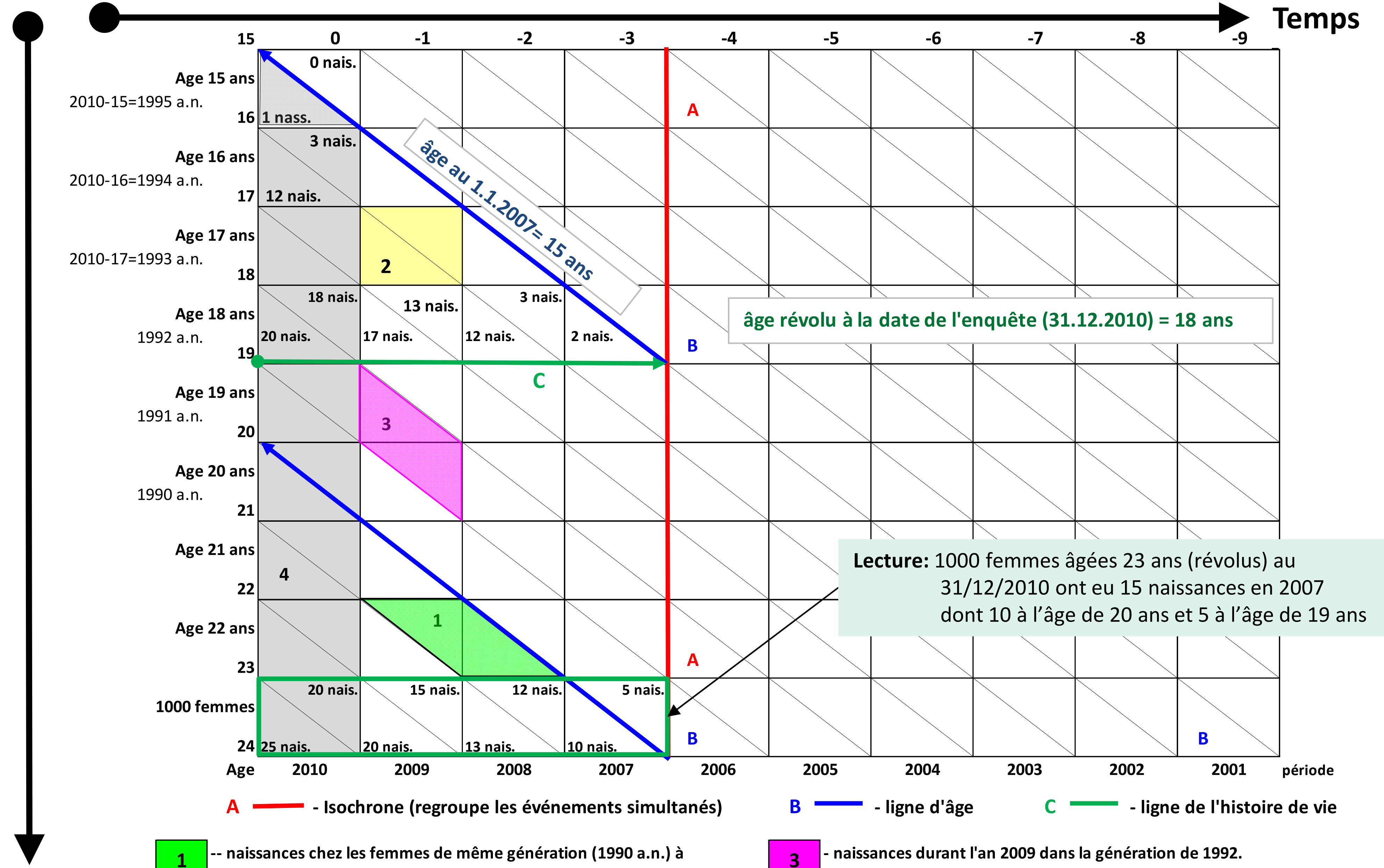
G – génération (date de naissance)

<u>Déjà vu dans</u>

Knapp, Georg (1874)- *Theorie des Bevölkerungswechsels*, Brunswick.

Becker, Karl(1874)- *Zur Berechnung von Sterbetafeln an die Bevölkerungsstatistik zu Stellende Anforderungen*, Berlin.

Diagramme de Lexis appliqué à la tabulation « histoire de vie » des données d'une enquête rétrospective



A - Isochrone (regroupe les événements simultanés) B - ligne d'âge C - ligne de l'histoire de vie

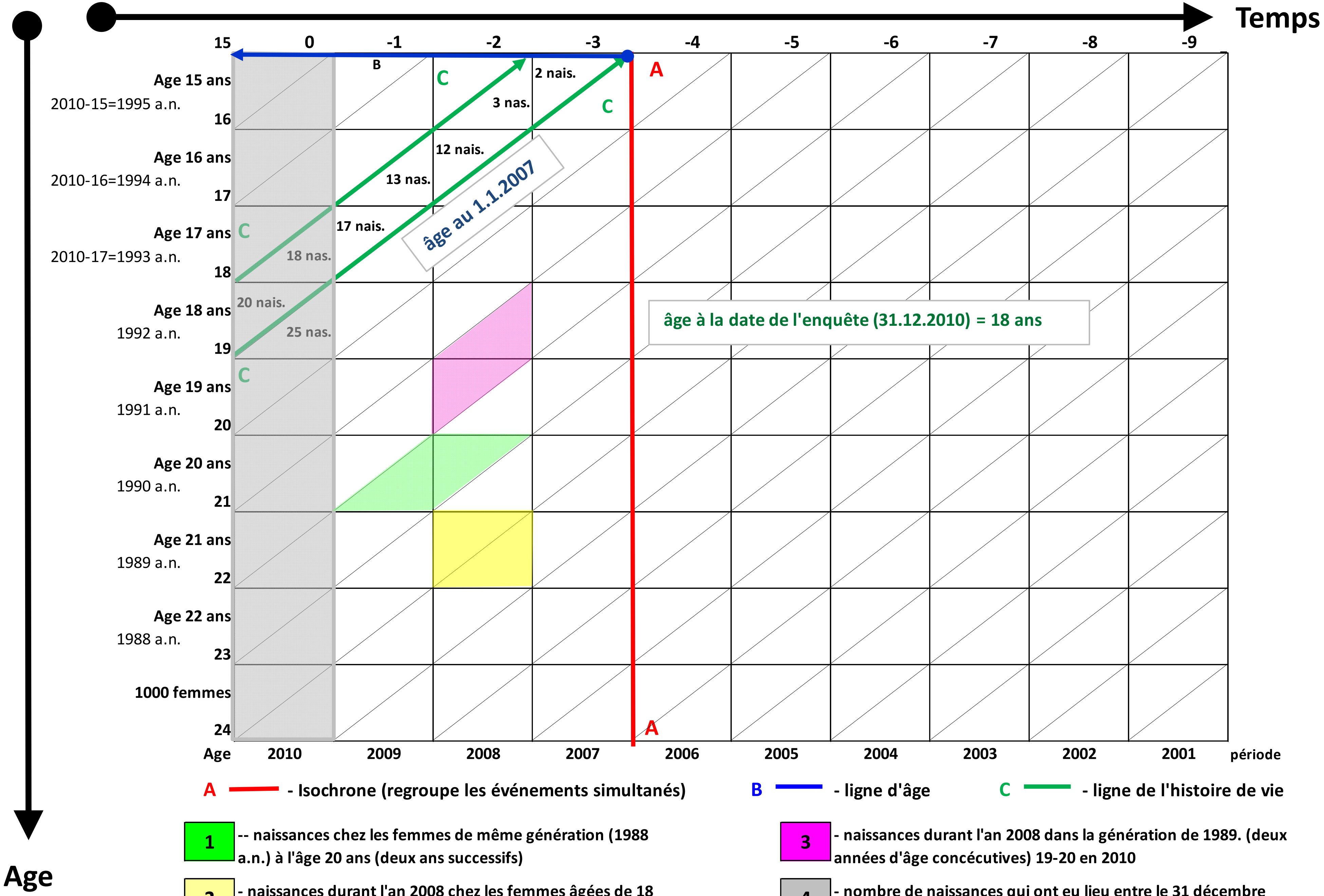
1 -- naissances chez les femmes de même génération (1990 a.n.) à l'âge 22 ans (deux ans successifs)

3 - naissances durant l'an 2009 dans la génération de 1992. (deux années d'âge concécutives) 19-20 ans au 31.12.2010

2 - naissances durant l'an 2009 chez les femmes âgées de 17 ans au 31.12.2010 (deux générations consécutives)

4 - nombre de naissances qui ont eu lieu entre le 1 janvier 2010 et le 1 janvier 2011 (durant l'an 2010)

Diagramme de Lexis appliqué à la tabulation des données transposées d'une enquête rétrospective



A - Isochrone (regroupe les événements simultanés)

B - ligne d'âge

C - ligne de l'histoire de vie

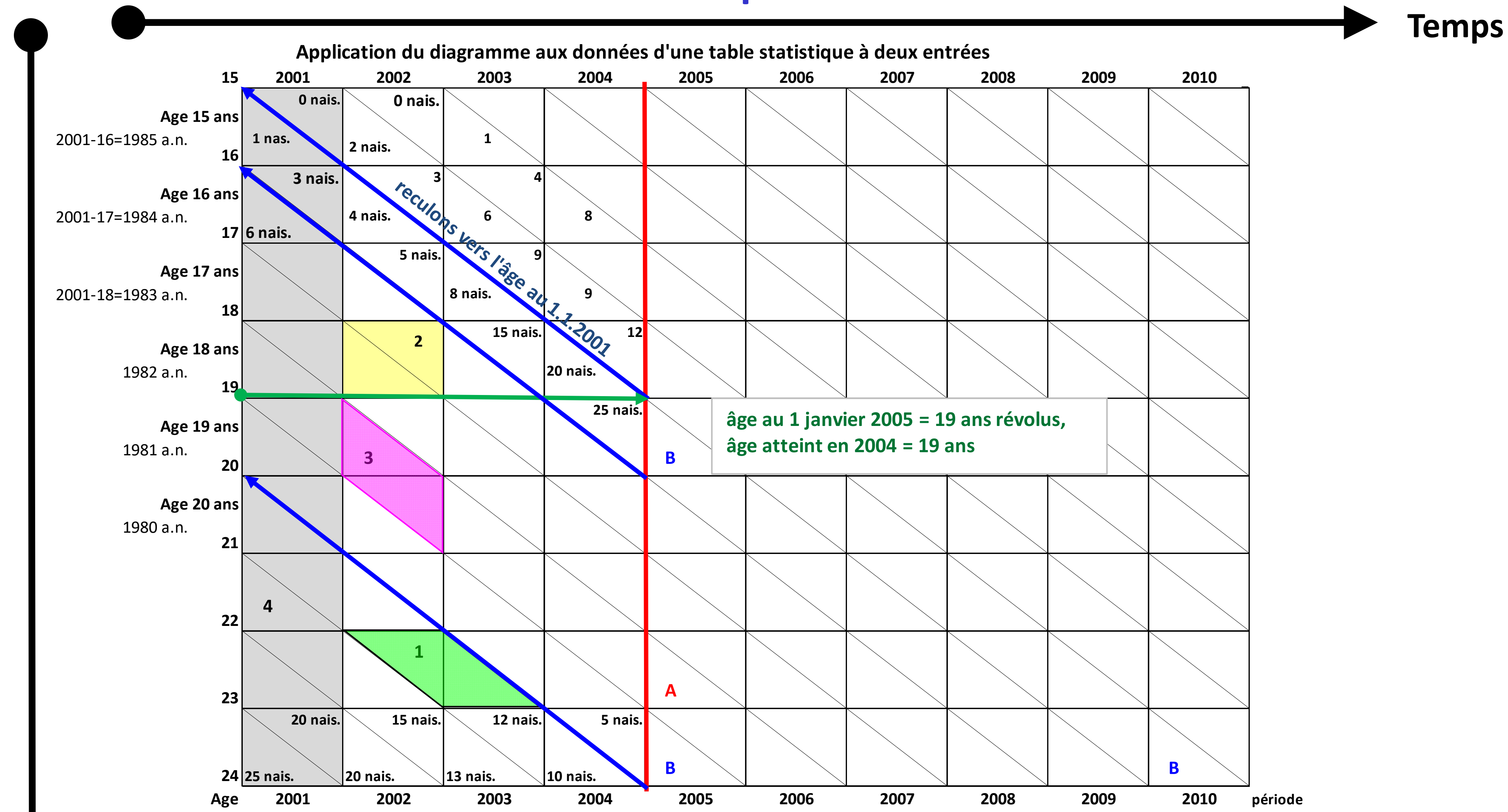
1 -- naissances chez les femmes de même génération (1988 a.n.) à l'âge 20 ans (deux ans successifs)

2 - naissances durant l'an 2008 chez les femmes âgées de 18 ans au 31.12.2010 (deux générations consécutives)

3 - naissances durant l'an 2008 dans la génération de 1989. (deux années d'âge concécutives) 19-20 en 2010

4 - nombre de naissances qui ont eu lieu entre le 31 décembre 2009 et le 1 janvier 2010 (durant l'an 2010)

Diagramme de Lexis appliqué à la tabulation classique des données de la statistique courantes



A - Isochrone (regroupe les événements simultanés)

B - ligne d'âge

1 -- naissances chez les femmes de même génération (1980 a.n.) à l'âge de 20 ans révolus (deux ans successifs)

2 - naissances durant l'an 2002 chez les femmes âgées de 18 ans révolus (deux générations consécutives)

3 - naissances durant l'an 2002 dans la génération 1982 (2002-20=1982 deux années d'âge consécutives) 19 et 20 ans en 2002 (âge atteint =20 ans)

4 - nombre de naissances enregistrées en 2001 (entre le 1 janvier 2001 et le 1 janvier 2002)

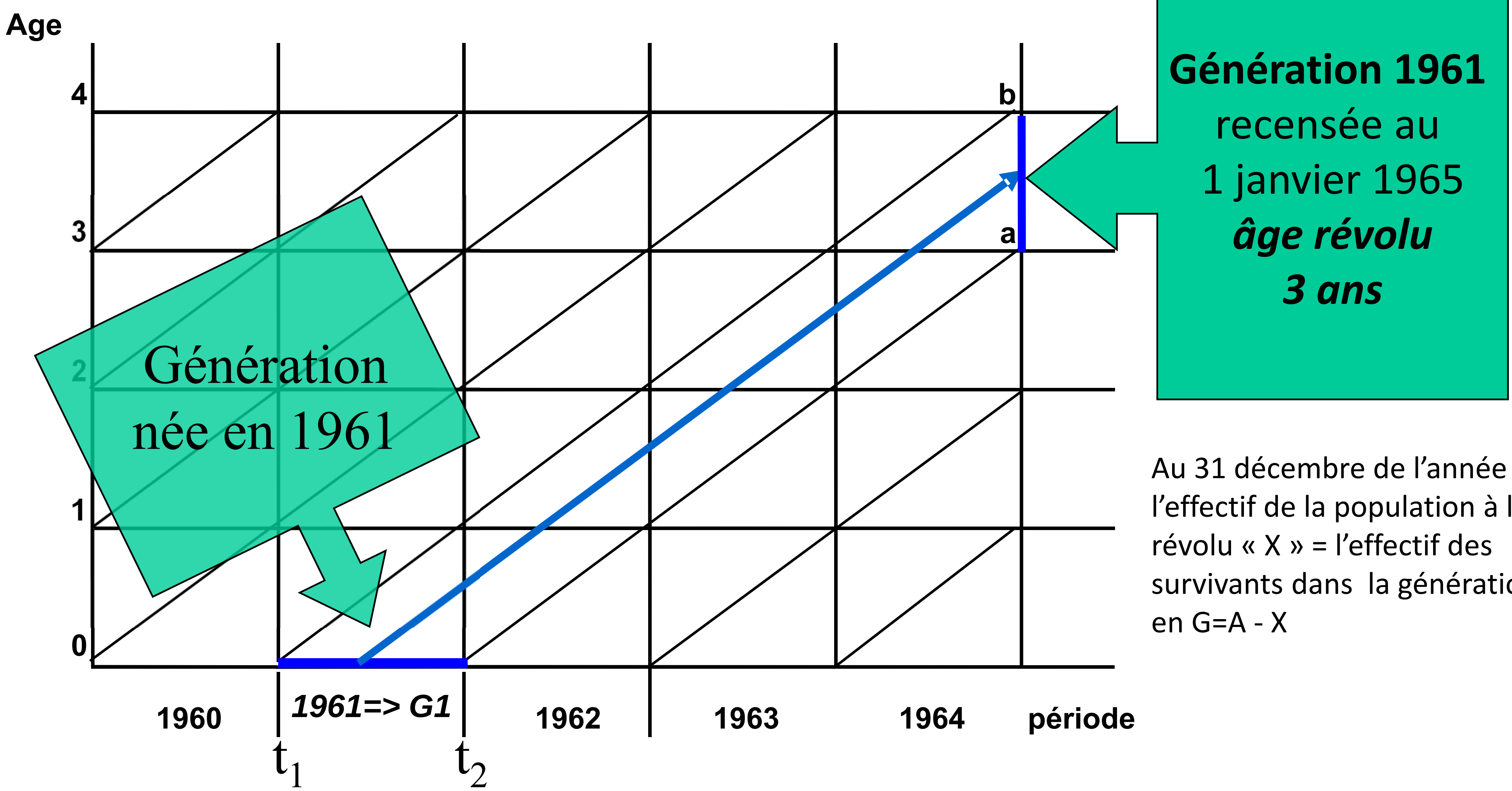
1. Repérage des populations

- Générations
- Age révolus
- Population moyenne

<u>*Génération :*</u>

L'ensemble des individus qui sont nés entre les deux dates exactes. Autrement dit « la génération » est un ensemble des individus dont les dates de naissance (DN) sont conforme à l'inégalité : $t_1 < DN < t_2$

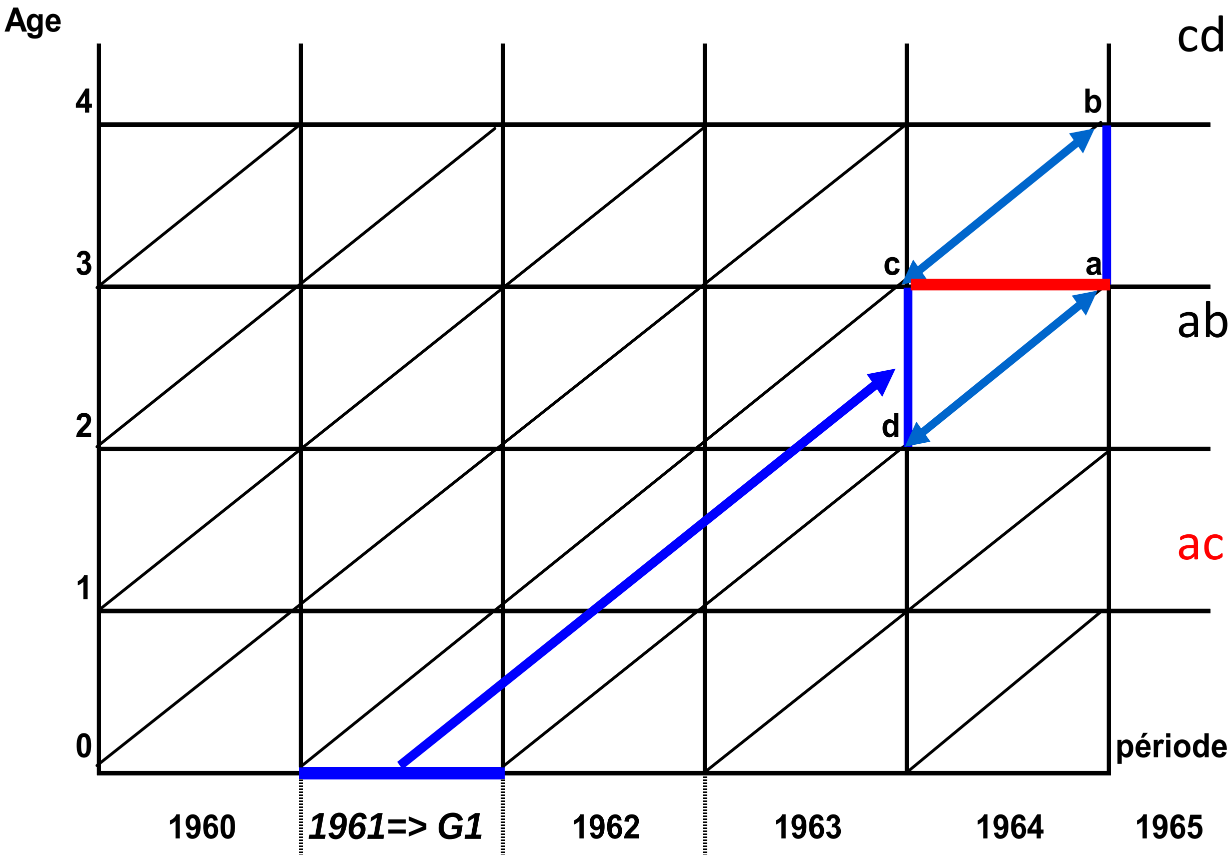
Comme l'année civile est une unité usuelle de temps dans la démographie, on l'utilise le plus souvent comme le critère de la génération.



Au 31 décembre de l'année « A » l'effectif de la population à l'âge révolu « X » = l'effectif des survivants dans la génération née en G=A - X

Exemple 1 : les survivants à l'âge exact

On s'intéresse à l'effet de cohorte et on cherche la population exposée dans la dimension génération-période une telle statistique n'existe nulle part, mais on peut l'estimer



ac ≈ 0,5 x (ab + cd) x 1 →

→ le nombre d'années vécues par la génération G1 durant l'an 1964

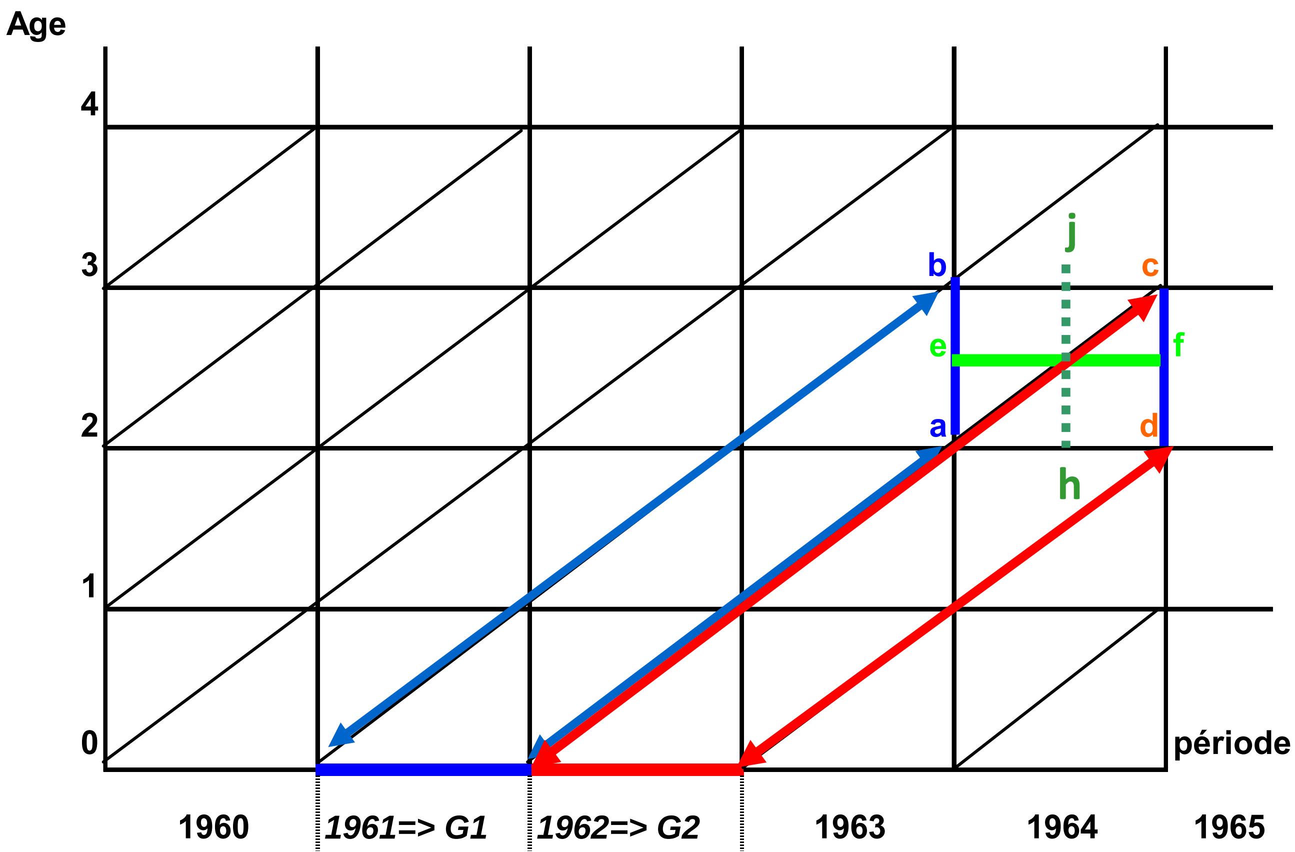
cd – ***ligne de date*** (de l'état) =
= l'effectif (survivants) de la génération G1 au 1 janvier 1964 (âge révolu)

ab – ***ligne de date*** =
= l'effectif de la génération G1 au 1 janvier 1965 **(31 décembre 1964)**

ac – ***ligne d'événement***
(par ex. : anniversaire ou naissance) → pour une période) =
= l'effectif moyen de la génération G1 pour l'an 1964 = population qui a atteint l'âge 3 ans en 1964
(cette ligne groupe des événements qui ont eu lieu durant une période déterminée, et non simultanément)

Exemple 2 : Population dite « moyenne »

ou plus exactement la population au milieu d'une période



On s'intéresse à l'effet de l'âge (et période)

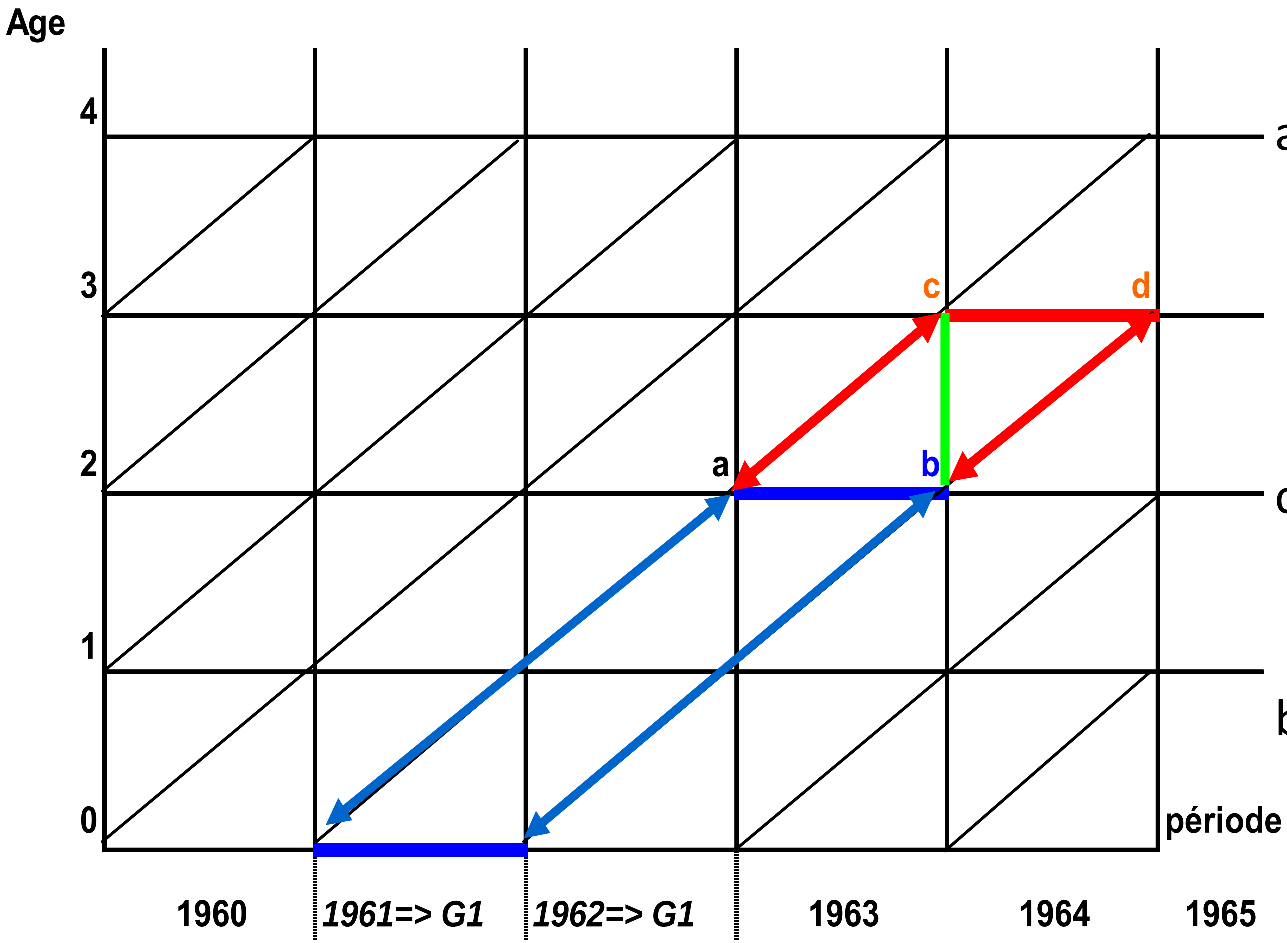
ab – effectif de la génération G1 au 1 janvier 1964 (âge 2 ans révolus)

cd – effectif de la génération G2 au 1 janvier 1965 (âge 2 ans révolus)

ef ≈ jh ≈ 0,5 (ab + cd) → *la population à l'âge atteint 2,5 en 1964 : une partie de la G1 + une partie de la G2 (moitié , moitié) ;*

la population « moyenne » à l'âge 2 ans révolu en 1964 : le pointillé vert foncé (jh) ≡ un équivalent statistique de la population à l'âge 2 ans révolu au 30 juin 1964 ;

Exemple 3 : Population du moment ou une population recensée



ab – l'effectif de la génération **G1**, qui a atteint l'âge de *2 ans* en 1963

(on dit: l'âge en différence des millésimes 1963 – 1961 =2)

cd – l'effectif de la génération **G1**, qui a atteint l'âge de *3 ans* en 1964

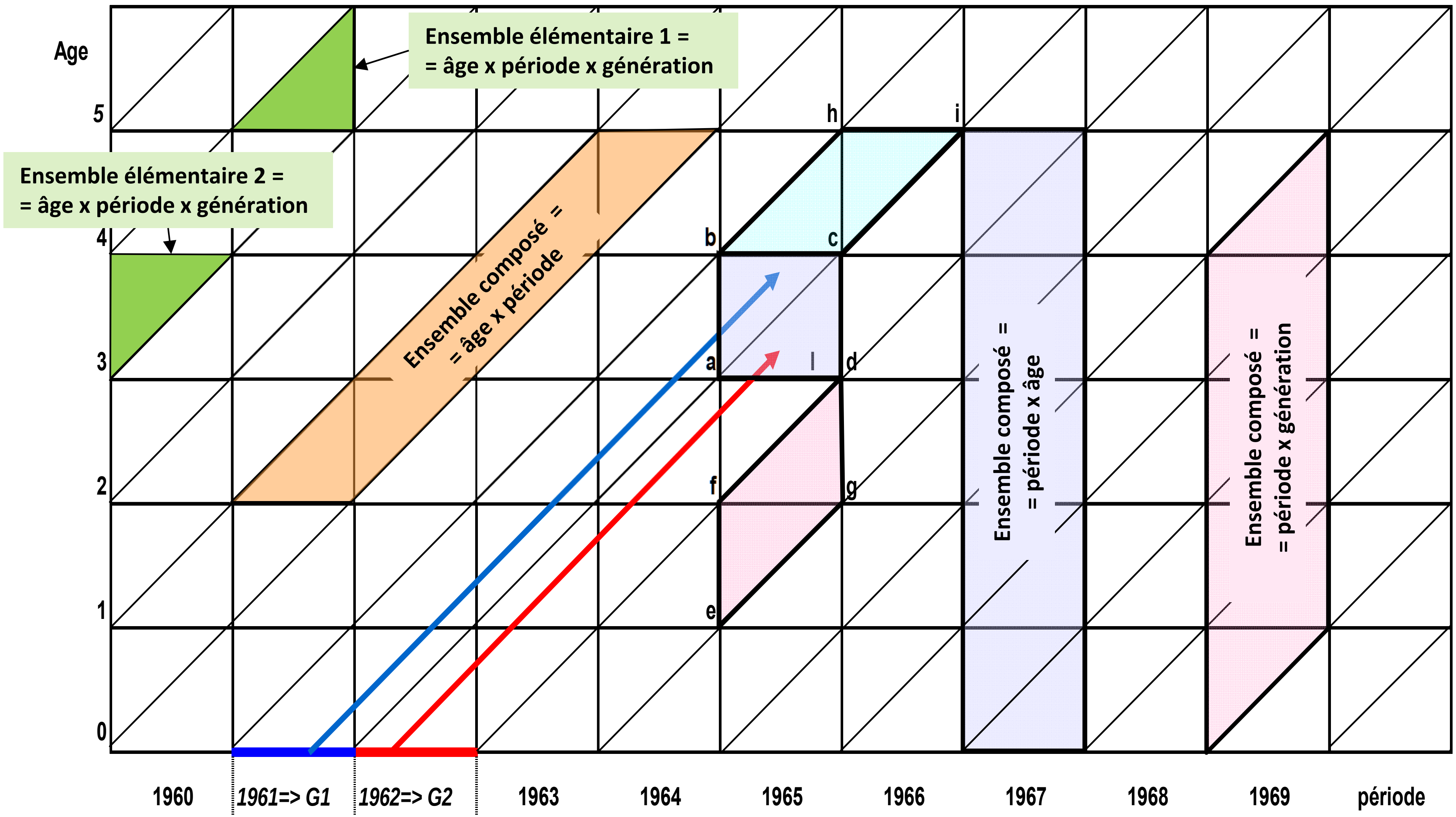
bc – l'effectif de la génération **G1** au 1 janvier 1964

cb ≈ 0,5 (ab + cd) → *la population « moyenne » d'une génération à l'âge révolu de 2 ans au 1 janvier 1964*

2. Repérage des événements démographiques

- Les ensembles d'événements démographiques (carré et parallélogrammes)
- Les ensembles élémentaires (triangles)

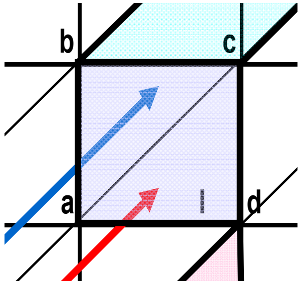
Localisation des événements démographiques sur le diagramme :
les ensembles annuels et pluriannuels



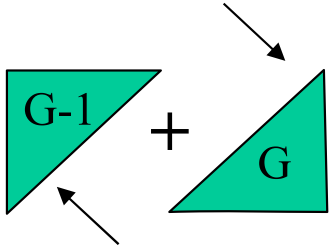
Carré ABCD identifie les événements selon l'âge et période produits par une population qui a vécu dans cet âge et durant cette période (≡ Nombre d'Années Vécus, NAV)

Trois types de classement annuel des événements démographiques

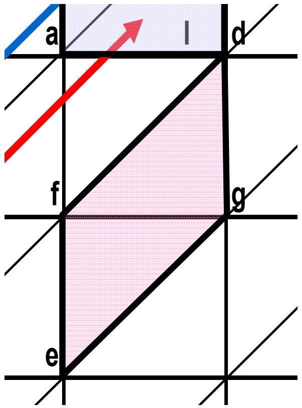
Forme | **Composition** | **Description**



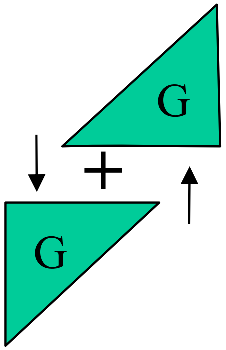
=



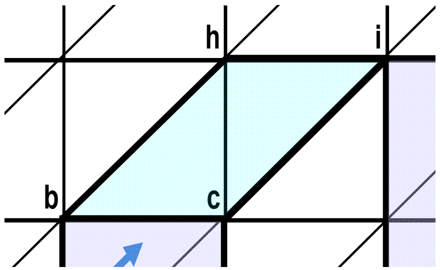
Type I:
(Age révolu) x (période)



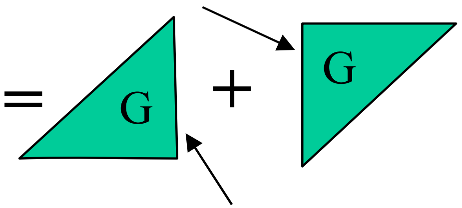
=



Type II:
(Génération) x (période)

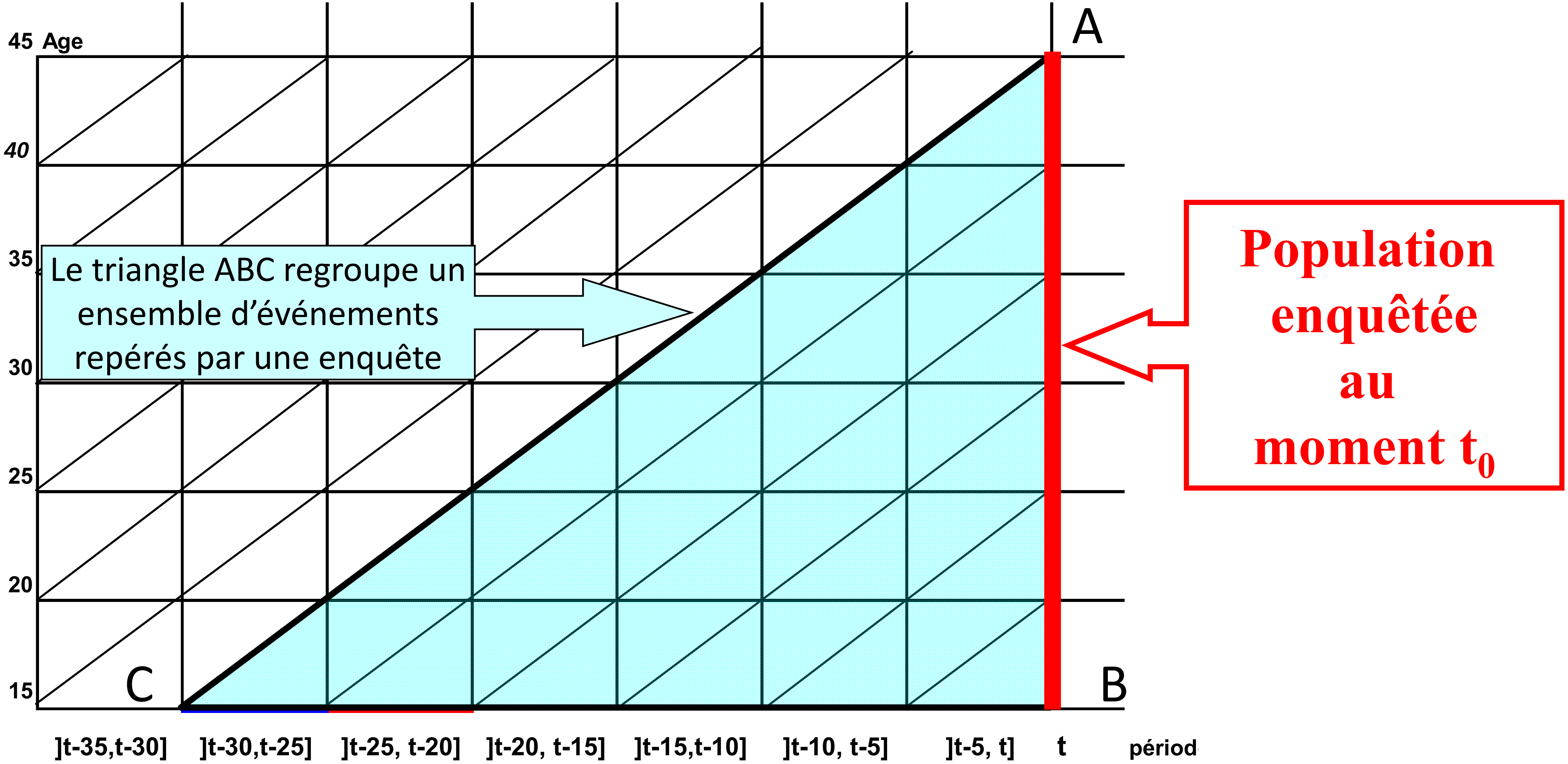


=



Type III:
(Age révolu) x (génération)

Ensemble des événements démographiques dans les enquêtes rétrospectives



Les défauts des enquêtes rétrospectives:

a) dans la dimension longitudinale : les événements repérés sont souvent les événements sélecteurs

b) dans la dimension transversale : cf. a) plus la perte des informations relatives aux âges élevés

Quelles informations faut-il chercher, quelles questions faut-il poser ?

- La meilleure solution est de poser les questions **sur les dates exactes de chaque événement** (jour, moi, année), en commençant par la date de naissance du répondant ou d'une personne d'intérêt.
- Ces questions pourraient être complétées pas des questions sur l'âge du répondant au moment de l'événement (surtout si ce dernier ne se souvient pas des dates, ou s'en trompe).
- Si les dates sont connues, il n'est pas inutile de créer telles variables de travail comme :
  - *L'âge au moment de l'événement* (en années révolue, ou éventuellement en mois) pour pouvoir préciser la durée exacte de la période de risque
  - *L'âge atteint dans l'année* (ou l'âge à la fin de l'année) **en différence des millésimes** (l'année d'événement moins l'année de naissance) qui utile pour l'analyse longitudinale (des générations).
  - *Une durée d'intervalles en unité de temps plus petites*, i.e. en mois, en semaines, en jours. *(Si votre logiciel d'analyse statistique permet de travailler avec les variables au format de « date », vous pouvez en négliger)*

Qu'est-ce qu'on peut faire avec les informations sur les dates des événements ?

- Estimer l'effectif de la population dans le passé (pour une période antérieure d'une enquête exhaustive ou sur échantillon): cette procédure est nécessaire pour balancer les résultats de deux recensements successifs).
- Estimer le niveau et la structure par âge de la mortalité pour une période antérieure d'une enquête
  (**attention !** Il faut considérer telles estimations avec beaucoup de précaution).
- Estimer le niveau et l'évolution de la mortalité infantile.
  (Généralement, telles estimations de la mortalité infantile sont plus exactes que celles de la mortalité aux âges adulte, et surtout que celles aux âges élevés et des personnes seules)
- Estimer la fécondité pour une période antérieure d'une enquête avec des méthodes relativement simples comme P/F, e.g. , ainsi que avec celles plus sophistiquées basées sur l'analyse des biographies individuelles.

3. Calculs des taux et des quotients à l'aide du diagramme de Lexis-Pressat

- Calculs des taux de différents types
- Conversion des taux
- Calculs des quotients de différents types
- Conversion des quotients
- Taux/quotients de « première » et de « seconde » catégorie

Rappel

Taux en démographie

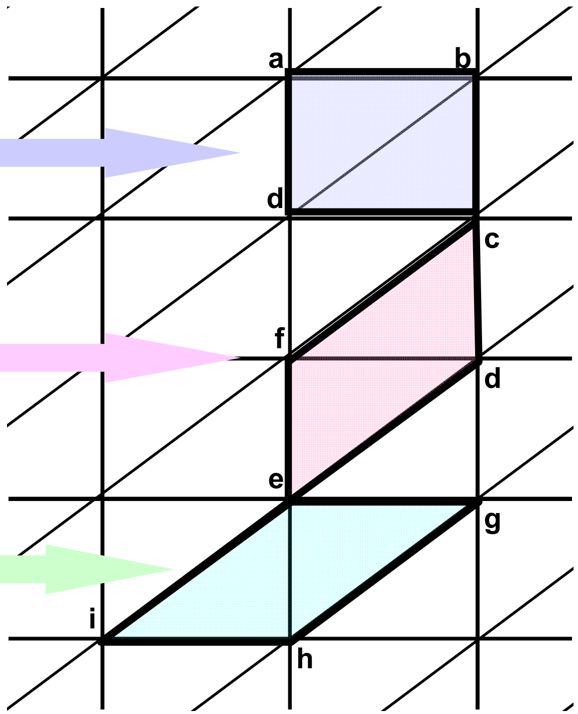
$$\text{Taux} = \frac{\text{Nombre d'occurences}^{1)}}{\text{Nombre de personnes} - \text{années d'exposition au risque}}$$

Trois types de classement des événements démographiques

Type I — **(Âge) X (Période)**

Type II — **(Génération) X (Période)**

Type III — **(Âge) X (Génération)**



1) Occurrence : « Chaque apparition explicite d'un élément dans un énoncé. » © Hachette Livre, 1998

Taux par âge de type I (âge-période)*

Exemple: mortalité

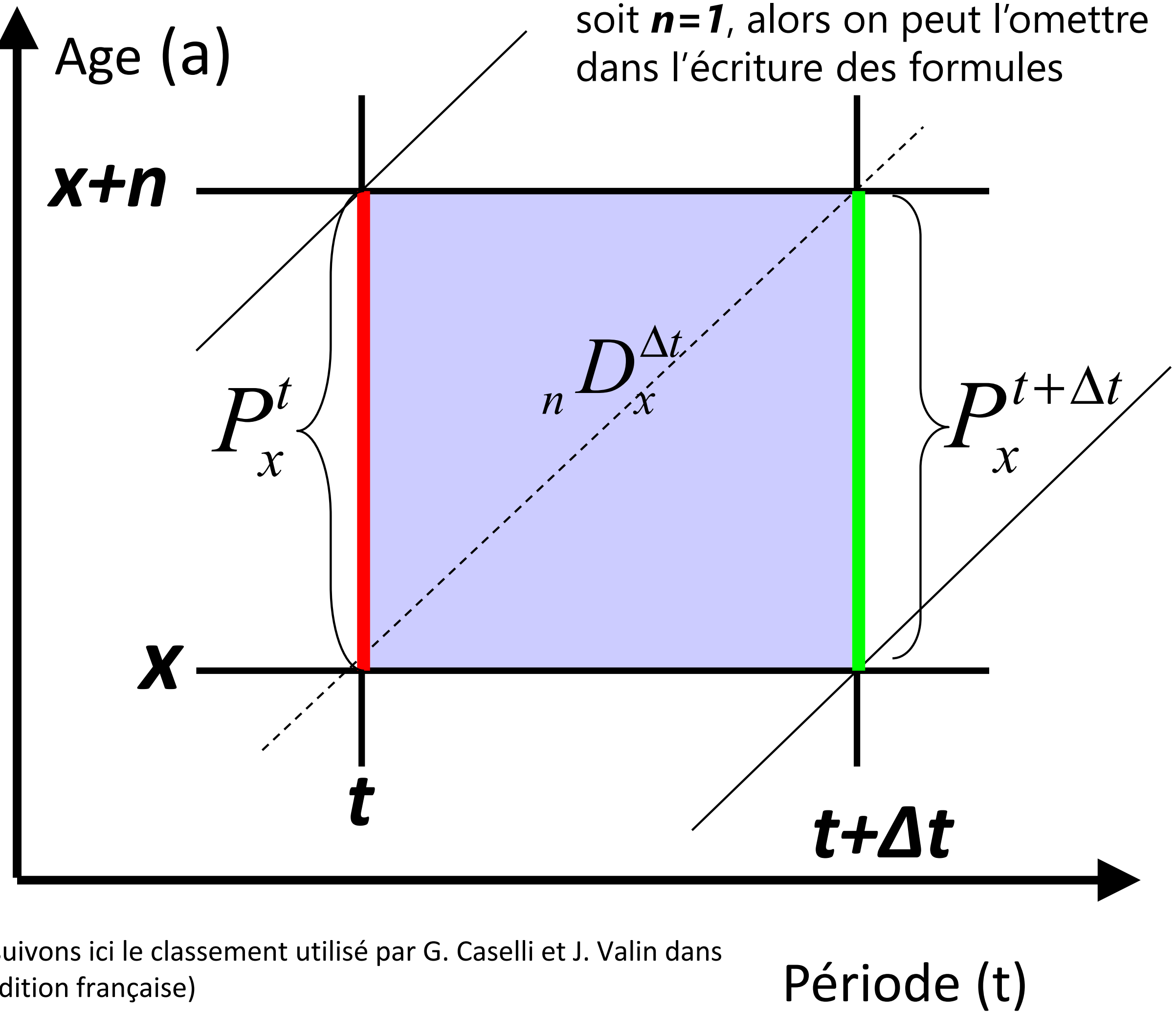
Ce taux (de mortalité par âge) mesure la mortalité moyenne pour une période et dans un intervalle l'âge (ce dernier inclut deux ou plusieurs générations consécutives)

$$ {}_n m_x^I = \frac{{}_n D_x^{\Delta t}}{0,5 \cdot (P_x^t + P_x^{t+\Delta t}) \cdot \Delta t} $$

${}_n D_x^{\Delta t}$ – nombre de décès à l'âge « ***x*** » révolu (entre âge exact ***x*** et ***x+n***) durant une période « ***t*** » (entre moment ***t*** et ***t+Δt***)

P_x^t – population à l'âge « ***x*** » révolu (entre âge exact ***x*** et ***x+n***) au moment « ***t*** ».

$P_x^{t+\Delta t}$ – population à l'âge « ***a*** » révolu (entre âge exact ***a*** et ***a+1***) au moment « ***t+1*** ».



* - cette numérotation (ordre) de types de taux est arbitraire, nous suivons ici le classement utilisé par G. Caselli et J. Valin dans *Démographie: analyse et synthèse*, 2001, vol.1, ch.6 (p.103 de l'édition française)

Taux de type II (dit « perspectif » ou *période-génération*)

Ce taux mesure la mortalité moyenne d'une ou plusieurs générations consécutives dans un intervalle de temps

$$m^{II}_{G(t-x)} = \frac{D^{G(t-x)}_{x-1,x+1}}{0,5 \cdot (P^{t}_{x-1} + P^{t+1}_{x}) \cdot t}$$

Soit – ***x*** âge atteint[1)] dans l'année (période) ***t*,** tel que ***t* – *x* = *g*** (année/période de naissance)

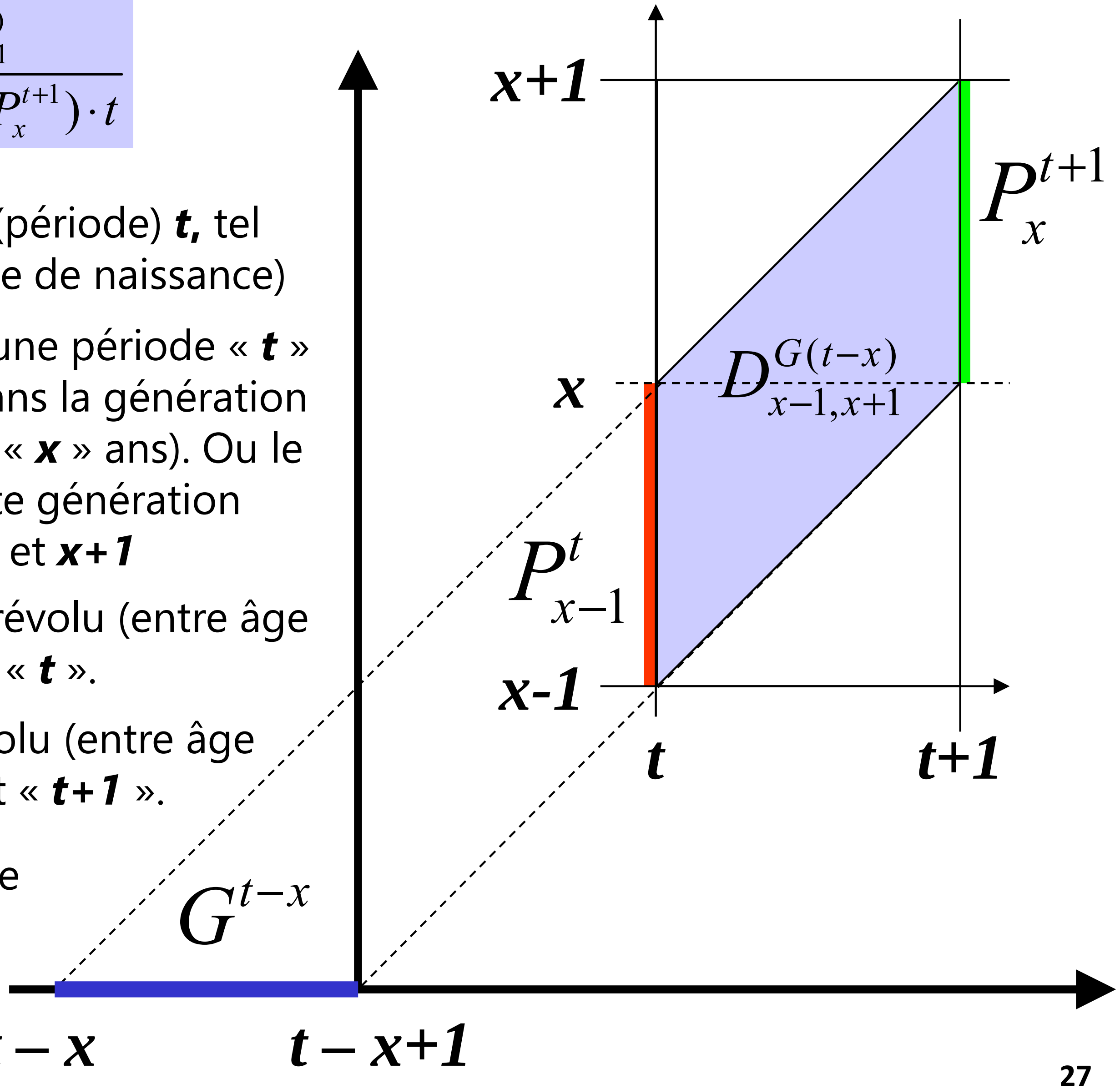
$D^{G(t-x)}_{x-1,x+1}$ – nombre de décès pendant une période « ***t*** » (entre moment ***t*** et ***t+1***) dans la génération née entre ***t-x*** et ***t-x+1*** (il a « ***x*** » ans). Ou le nombre de décès dans cette génération entre deux âges exacts ***x-1*** et ***x+1***

P^{t}_{x-1} – population à l'âge « ***x-1*** » révolu (entre âge exact ***x-1*** et ***x***) au moment « ***t*** ».

P^{t+1}_{x} – population à l'âge « ***x*** » révolu (entre âge exact ***x*** et ***x+1***) au moment « ***t+1*** ».

G^{t-x} – une génération de naissance durant la période « ***t – 1*** ».

Exemple: mortalité



Note

1) On parle aussi de l'âge à la fin de l'année (période)

Taux de type III (*âge-génération*)

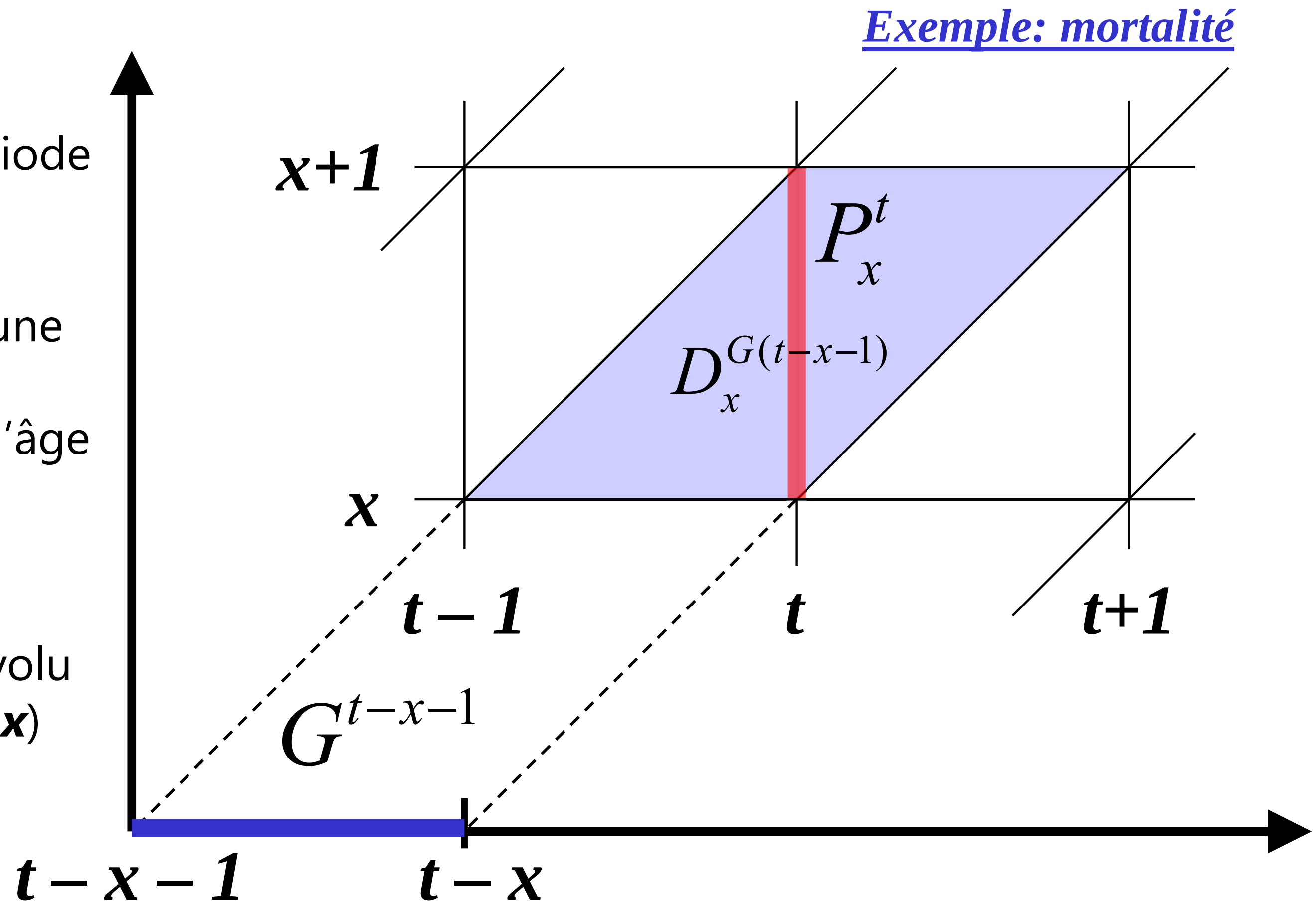
bon pour quantifier la densité moyenne de la probabilité d'un événement entre deux âges exacts dans une ou plusieurs générations (relevant du temps individuel)

$$m_x^{III} = \frac{D_x^{G(t-x-1)}}{P_x^t}$$

Soit – âge atteint dans l'année (période) ***t***, tel que ***t – x – 1 = g*** (année/période de naissance)

$D_x^{G(t-x-1)}$ – nombre de décès dans une génération née entre le moment ***t-x-1*** et ***t-x***) à l'âge ***x*** ans révolus (entre ***x*** et ***x+1***)

P_x^t – population à l'âge ***x*** révolu (entre âge exact ***x-1*** et ***x***) au moment ***t*** .



Ce taux mesure s'applique aux tables démographiques (mortalité, fécondité etc.) sous l'hypothèse de la stationnarité

Données nécessaires:

Effectifs de la population à la fin d'une période (projection ?)

Taux de type I → Décès durant une période entre ***t*** et ***t+Δt*** classés par âge révolu + Population par âge révolu au début et à la fin d'une période

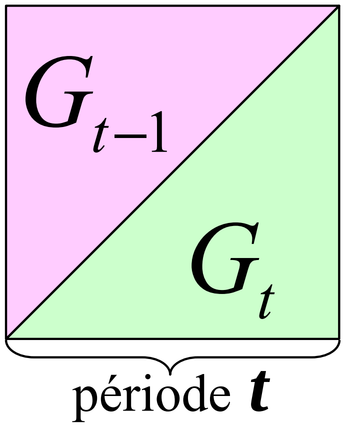
Taux de type II dit « perspectif » → Décès durant une période entre ***t*** et ***t+ Δt*** classés par années de naissance (et par âge) des décédés + Population par âge révolu au début et à la fin d'une période

Taux de type III → Décès durant une période entre ***t – Δt*** et ***t+ Δt*** classés par années de naissance (et par âge) des décédés + Population par âge révolu au milieu d'une période

Double classement des décès: par âge et par générations

Relations entre les taux de différentes catégories

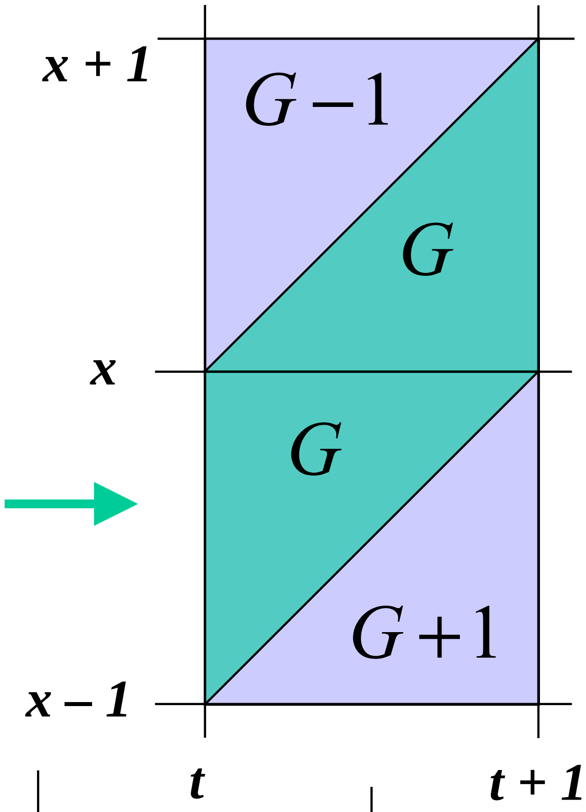
Le double classement des événements (par âge et par génération) permet de passer facilement d'un type de taux à l'autre.



Type I ⇒ Type II

$$m^{II}_{G(t-x)} = \alpha \cdot m^{I}_{x-1} + \beta \cdot m^{I}_{x}$$

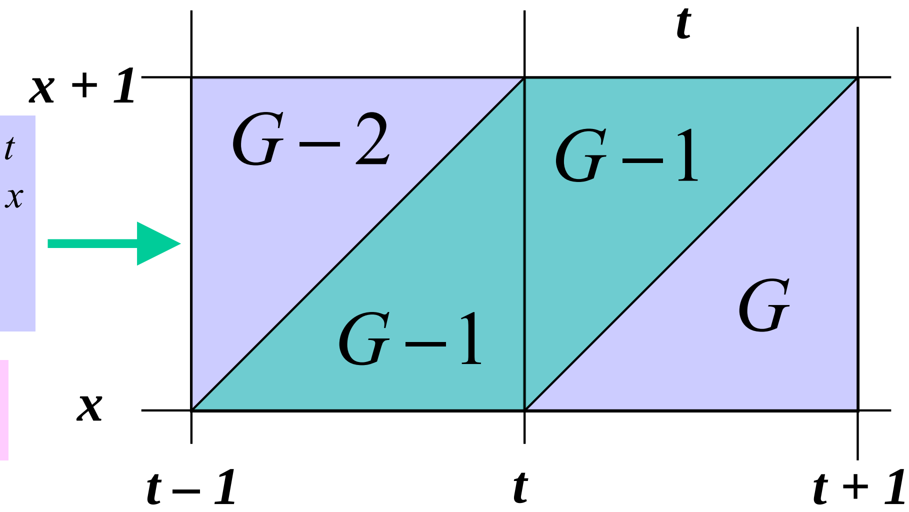
$$(\alpha + \beta = 1)$$



Type I ⇒ Type III

$$m^{III}_{G(t-x-1)} = \alpha \cdot m^{I\,t-1}_{x} + \beta \cdot m^{I\,t}_{x}$$

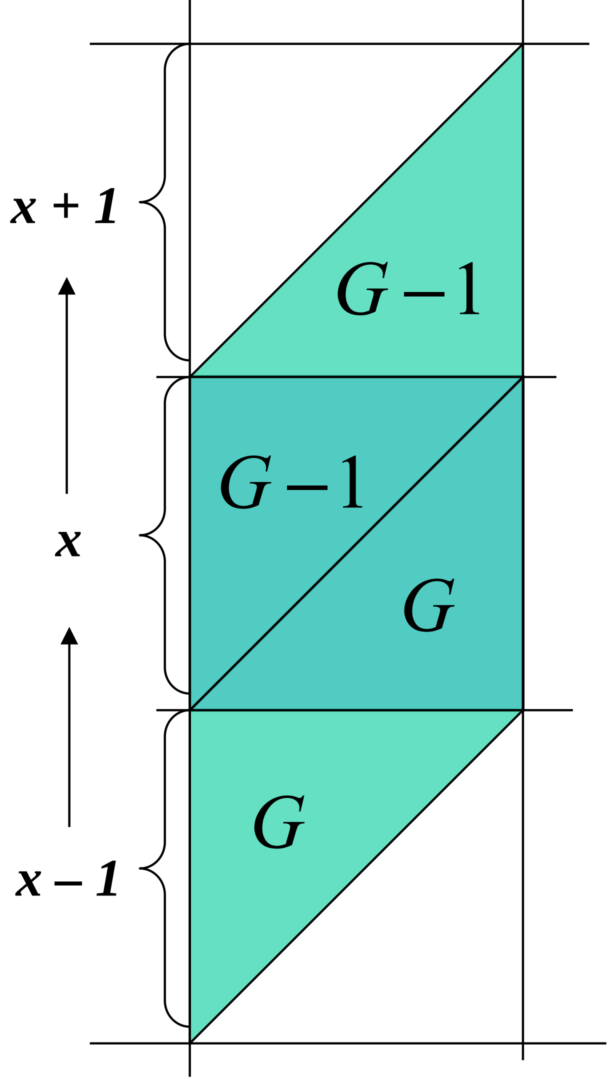
$$(\alpha + \beta = 1)$$



<u>Hypothèse de l'uniformité:</u>

$$\alpha = \beta = 0{,}5$$

Relations entre les taux de diverses catégories (suite)



$Type\ II \Rightarrow Type\ I$

$$m_x^{I} = \alpha \cdot m_x^{II} + \beta \cdot m_{x+1}^{II}$$

ou

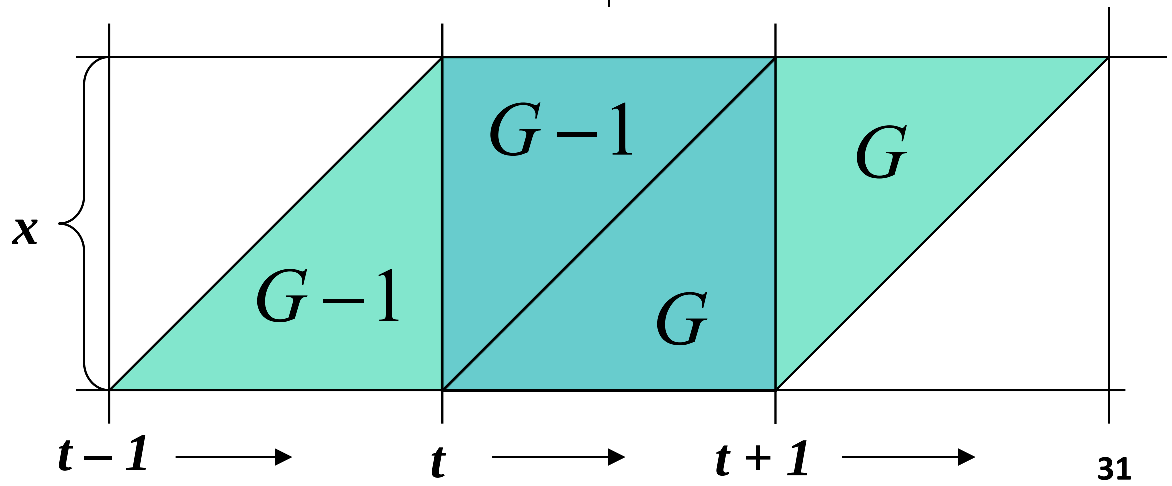
$\boldsymbol{x}$ – l’âge atteint dans l’année

$Type\ III \Rightarrow Type\ I$

$$m_x^{I\,t} = \alpha \cdot m_x^{III\,t} + \beta \cdot m_x^{III\,t+1}$$

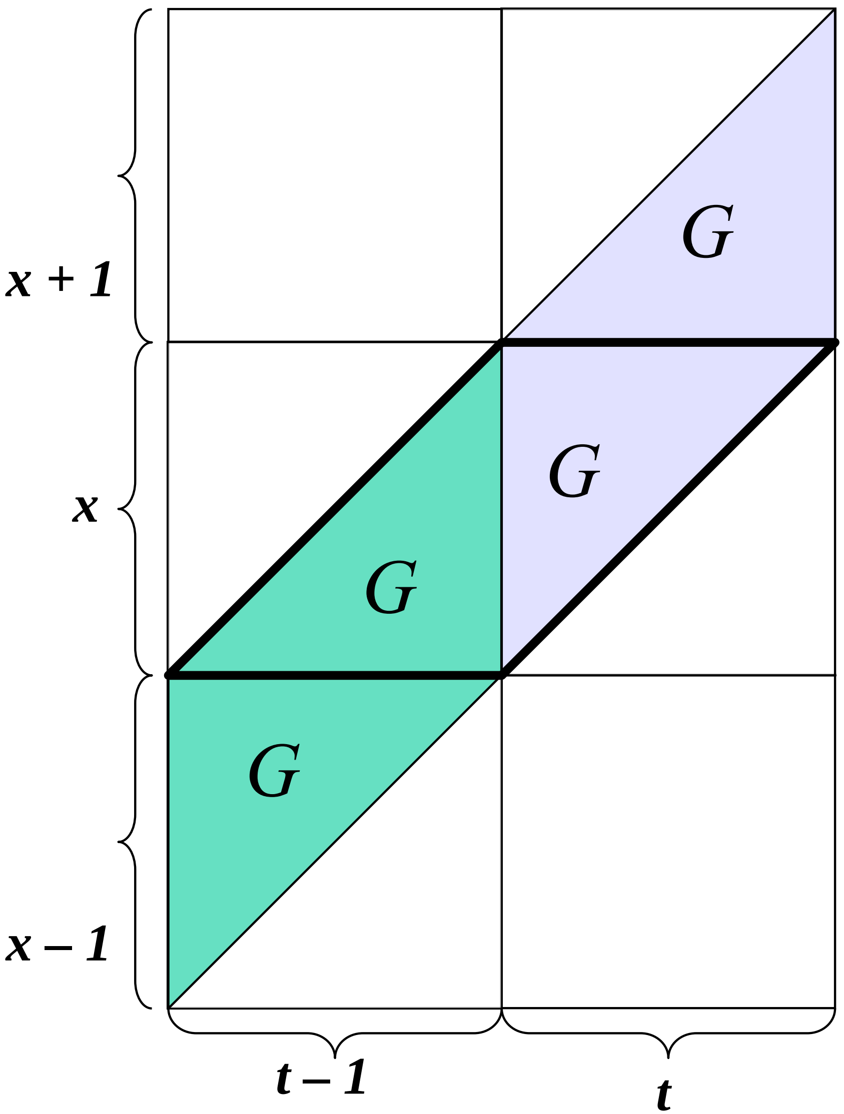
ou

$\boldsymbol{x}$ – l’âge révolu

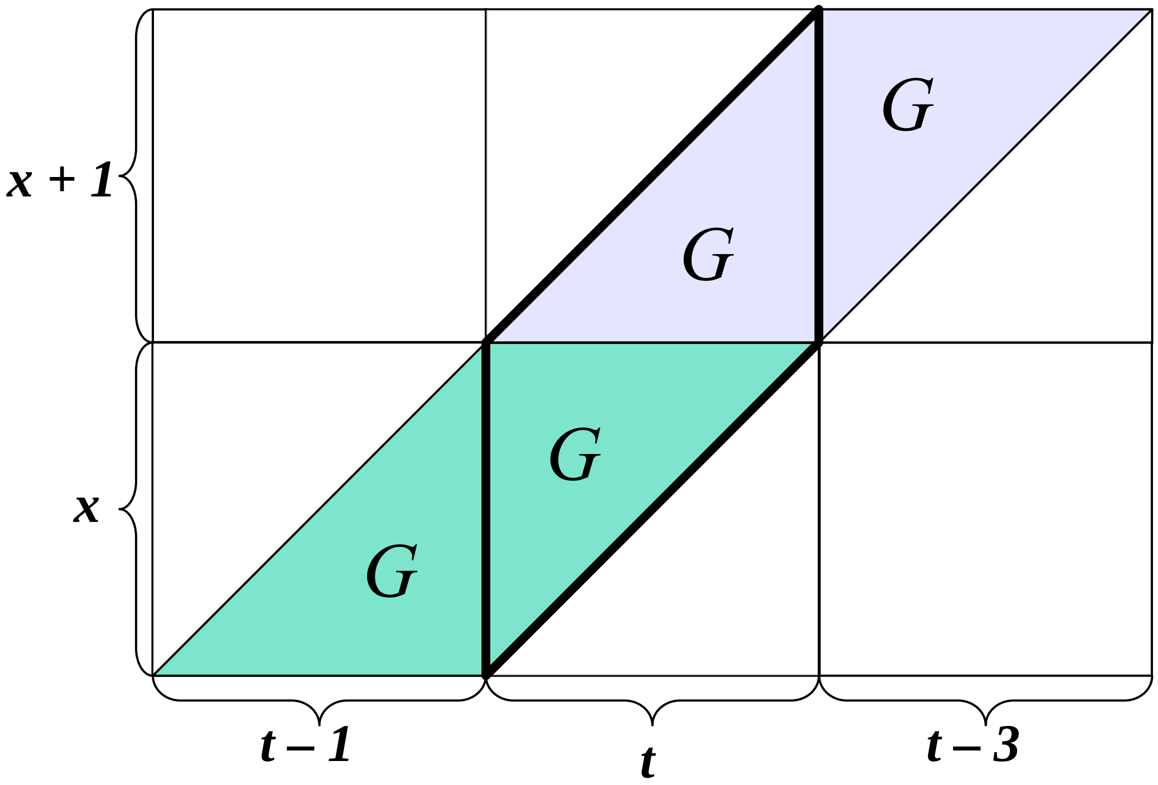


Relations entre les taux de diverses catégories (suite)

Type II ⇒ Type III



Type III ⇒ Type II



1. Quotient ou la probabilité d’un événement pour une personne appartenants à une catégorie

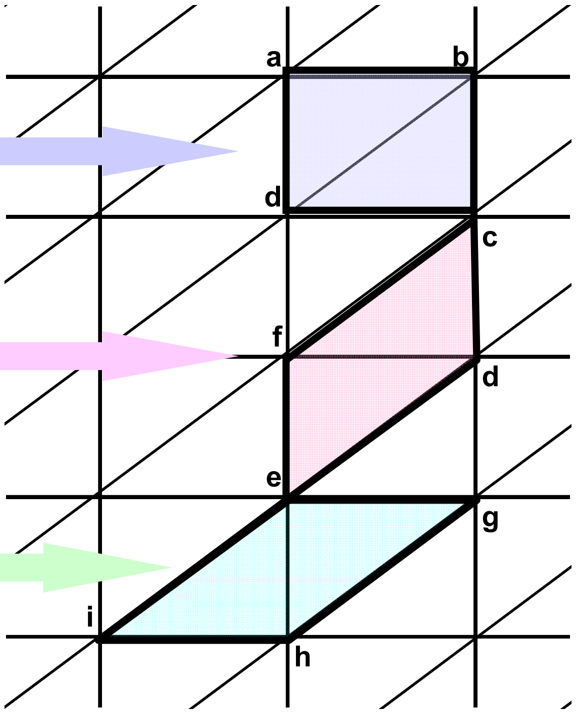
$$ {}_{N}Q_{X} = \frac{\text{Nombre d'occurences à l'âge X}}{\text{Population soumise au risque au début d'une période N}} $$

2. Trois types de classement des événements démographiques

Type I — (Âge) X (Période)

Type II — (Génération) X (Période)

Type III — (Âge) X (Génération)



Quotient « perspectif » (pour une génération et une période) associé à l'âge atteint dans l'année

$$q_{x-1,x+1}^{G(t-x)} = \frac{D_{x-1,x+1}^{G(t-x)}}{P_{x-1}^{t}}$$

Ce quotient mesure la probabilité de mourir durant un intervalle de temps (une année, e.g.) pour une génération (i.e. avant ou après X[e] anniversaire)

$D_{x-1,x+1}^{G(t-x)}$ – nombre de décès pendant une période « ***t*** » (entre moment ***t*** et ***t+1***) dans la génération née en ***t-x*** (différence des millésimes = il a « ***x*** » ans).

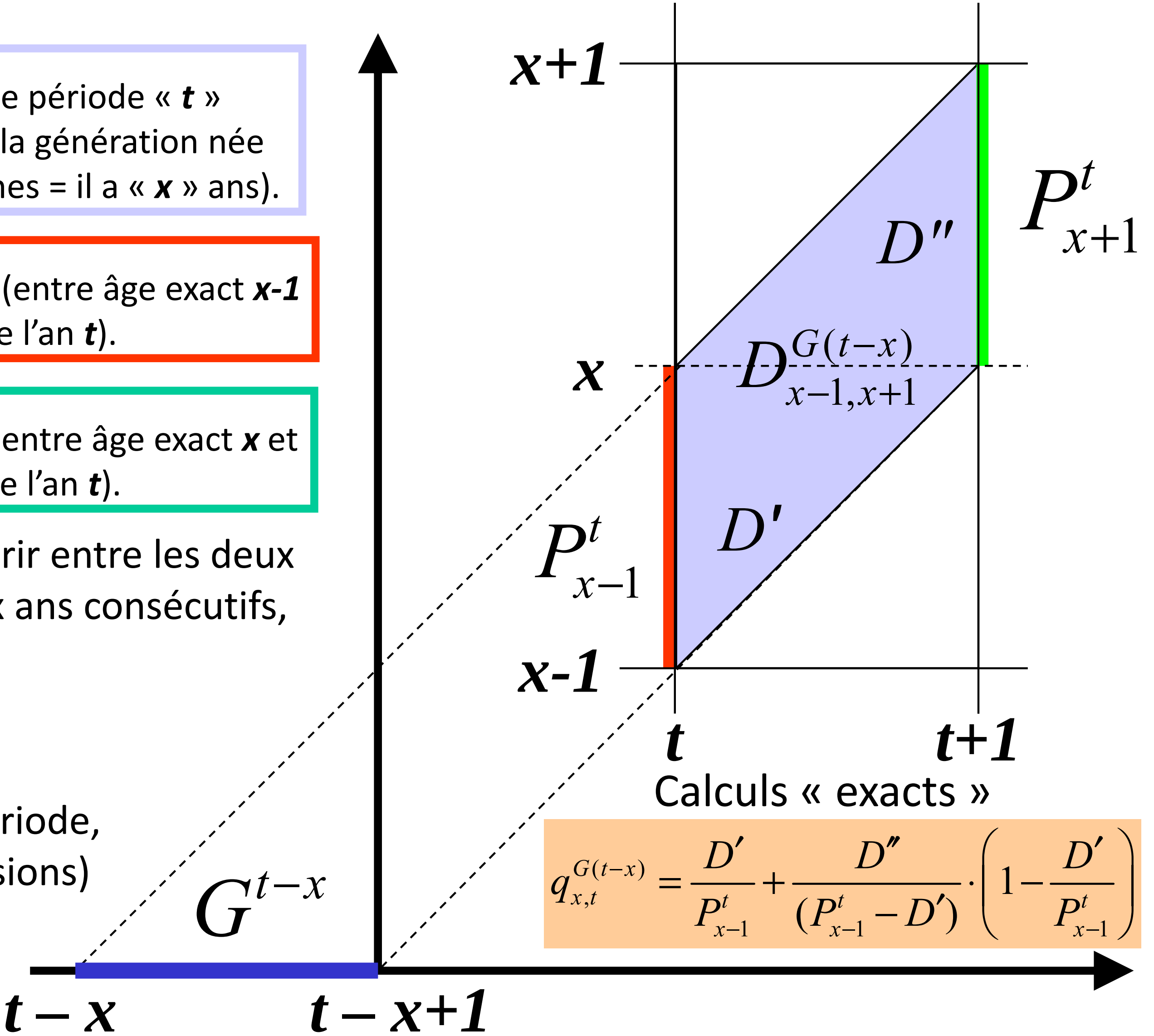
P_{x-1}^{t} – population à l'âge « ***x-1*** » révolu (entre âge exact ***x-1*** et ***x***) au moment « ***t*** » (début de l'an ***t***).

P_{x+1}^{t} – population à l'âge « ***x*** » révolu (entre âge exact ***x*** et ***x+1***) au moment « ***t+1*** » (fin de l'an ***t***).

On parle aussi de la probabilité de mourir entre les deux date (le 1 janvier et le 1 janvier de deux ans consécutifs, e.g.) pour une génération.

Ainsi $p_{x}^{G(t-x)} = 1 - q_{x}^{G(t-x)}$

la probabilité de survivre durant une période, qu'on utilise pour les projections (prévisions) démographiques



Calculs « exacts »

$$q_{x,t}^{G(t-x)} = \frac{D'}{P_{x-1}^{t}} + \frac{D''}{(P_{x-1}^{t} - D')} \cdot \left(1 - \frac{D'}{P_{x-1}^{t}}\right)$$

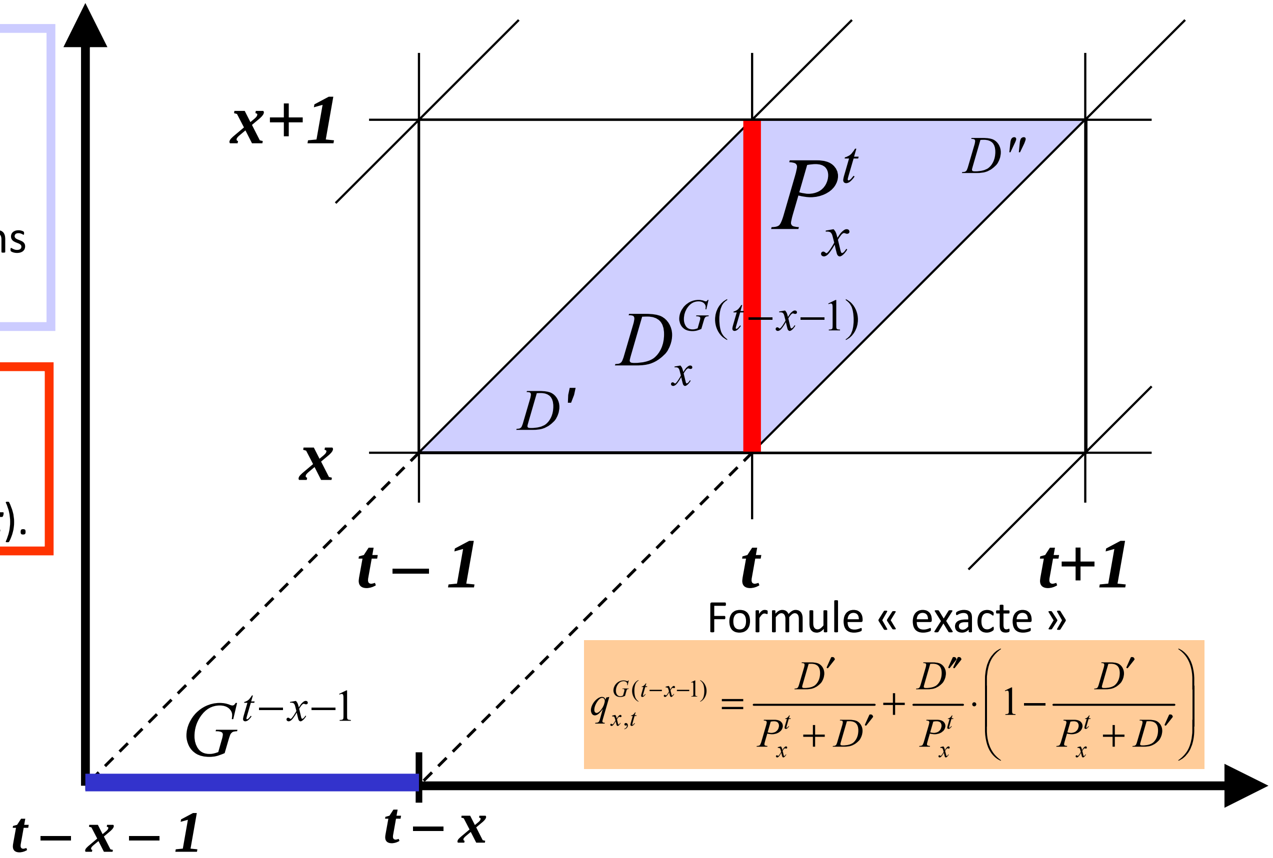
Quotient pour une génération entre deux anniversaires

$$q_x^{G(t-x-1)} = \frac{D_x^{G(t-x-1)}}{P_x^t + 0,5 \cdot D_x^{G(t-x-1)}}$$

Ce quotient mesure la probabilité d'un événement (e.g. un décès) pour une génération durant un intervalle d'âge

$D_x^{G(t-x-1)}$ – nombre de décès dans une génération née entre sur l'intervalle entre ***t-x-1*** et ***t-x*** (destinée avoir l'âge « ***x*** » ans révolus au début de l'an ***t***)

P_x^t – population à l'âge « ***x*** » révolu (entre âge exact ***x-1*** et ***x***) au moment « ***t*** » (au début de l'an ***t***).



Formule « exacte »

$$q_{x,t}^{G(t-x-1)} = \frac{D'}{P_x^t + D'} + \frac{D''}{P_x^t} \cdot \left(1 - \frac{D'}{P_x^t + D'}\right)$$

Puisque la statistique des anniversaires n'existe nulle part, on ne dispose jamais de l'effectif de population à l'âge exact. Il est nécessaire donc de l'estimer.

Cette estimation pourrait se faire à partir d'une hypothèse que les décès sont distribués uniformément dans l'intervalle d'âge (D'=D" → hypothèse d'uniformité). Une telle hypothèse n'est acceptable que ci les intervalles sont suffisamment courts. Sinon il faudra recourir aux hypothèses plus complexes (e.g. normalité de risque etc.), ou appliquer une formule « exacte », si les décès sont classés par génération et par âge (double classement)

Une probabilité de mourir associée à l'âge révolu et un intervalle de temps (âge x période)

Si nous supposons que la probabilité de mourir dans un triangle élémentaire ne change pas beaucoup durant deux années successives (***<u>hypothèse de stationnarité</u>***), on pourra donc en déduire que :

$$q_x = q'_x + q''_x - q'_x \cdot q''_x$$

On voit par ailleurs que sous cette hypothèse :

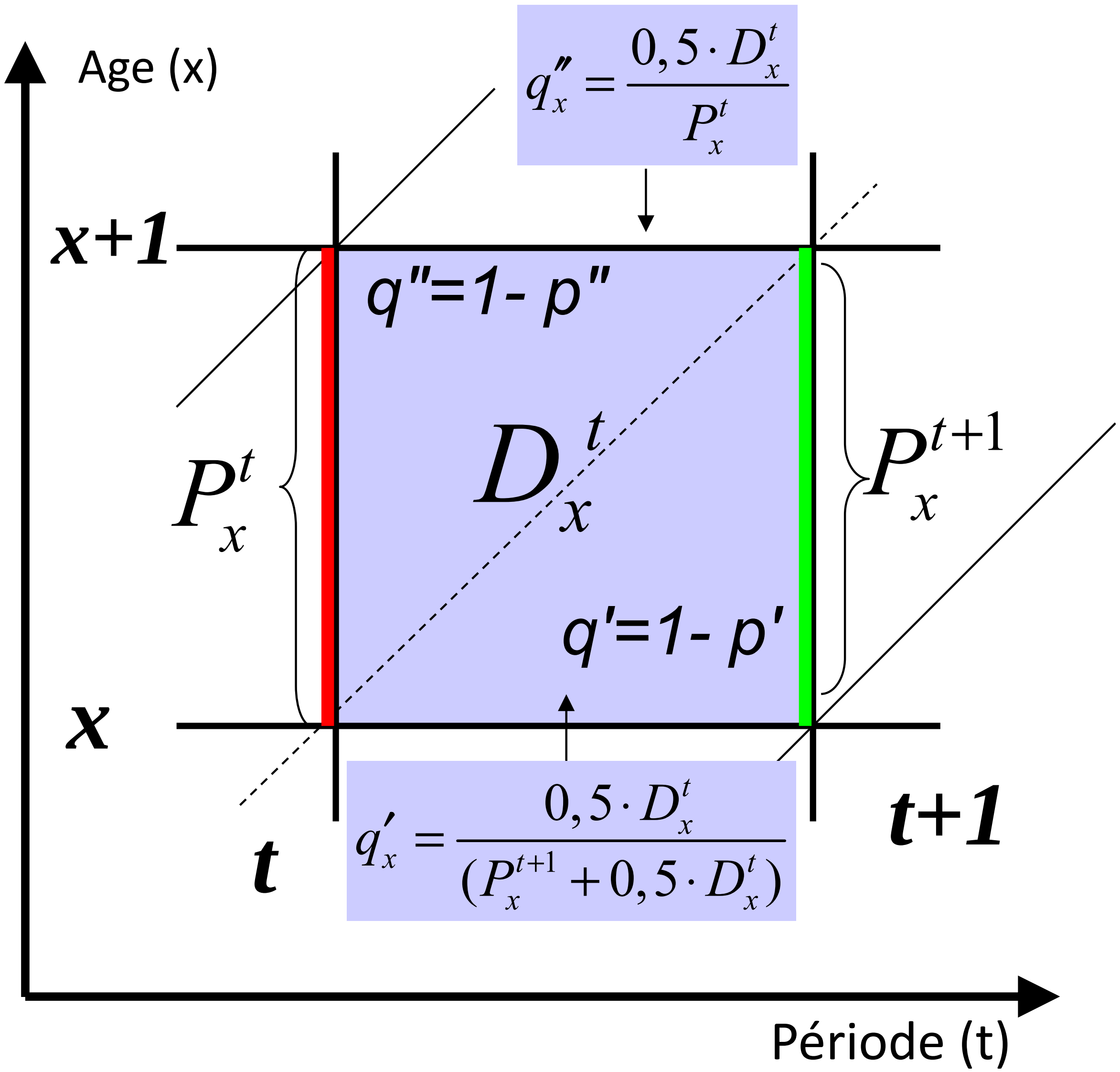
q′ ≈ ½ du quotient (de mortalité) de la génération P_x^t à l'âge *x* ou ≈ ½ du quotient (de mortalité) de la génération P_x^t pour l'année *t* ;

et

q″ ≈ ½ du quotient (de mortalité) de la génération P_x^{t+1} pour l'année *t ;*

Sachant que la valeur du produit (q′ x q″) est très faible, on peut le négliger et accepter la formule suivante :

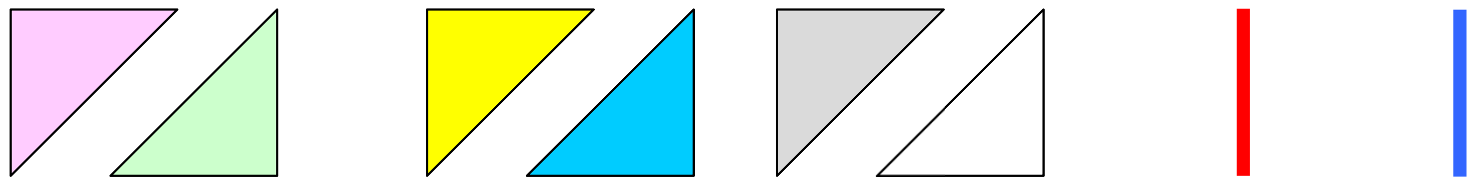
$$q_x = q'_x + q''_x$$



On recourt à des quotients de ce type (âge-période), si on s'intéresse à des situations très particulières (e.g. la mortalité infantile), ou si on étudie les périodes pluriannuelles

Variation des effectifs de générations: conséquences pour les taux:

une étude de l'effet d'âge (74 ans) et de période (an 1994)



À effectuer la transcription des données sur le diagramme de Lexis (version Pressat ou/et une autre)

	Tableau 76 INSEE 1993		1994		1995		Tableau 6 INSEE Population (1.1.xx)	
Age	**G-1**	**G**	**G-1**	**G**	**G-1**	**G**	**1994**	**1995**
69	3081	3300	3184	3205	3197	3279	224448	231630
70	3321	3600	3208	3240	3291	3457	219548	218205
71	3557	3619	3344	3498	3349	3425	213348	212922
72	3961	3801	3648	3719	3596	3613	213184	206391
73	2858	4183	4142	3878	3954	3925	210506	205693
74	2386	2318	**2889**	**4275**	4084	4280	***121312***	202140
75	2118	2393	2417	2275	3095	4646	***100716***	116196
76	2058	2163	2238	2438	2456	2376	***84490***	95863
77	2222	2259	2005	2137	2235	2667	***74641***	80091
78	4030	3326	2326	2242	2238	2376	***87657***	70442
79	4305	4294	4108	3249	2322	2345	***120679***	82035
80	4492	4560	4325	4171	4321	3444	113403	112397
81	4256	4738	4465	4400	4313	4308	104048	104663

Exemple de : Caselli G. et J.Vallin « Du repérage des événements dans le temps au diagramme de Lexis et auxcalcul des taux » Dans: Caselli G., J.Vallin et G.Wunsch *Démographie. Analyse et synthèse*. Vol.I « La dynamique des populations ». INED, Paris, 2001, p.102-109

Variation des effectifs de générations: conséquences pour les taux (suite)

G-1	G	Age
1924	1925	69
1923	1924	70
1922	1923	71
1921	1922	72
1920	1921	73
1919	**1920**	**74**
1918	**1919**	**75**
1917	1918	76
1916	1917	77
1915	1916	78
1914	1915	79
1913	1914	80
1912	1913	81

G = 1994 – âge

on observe six générations :

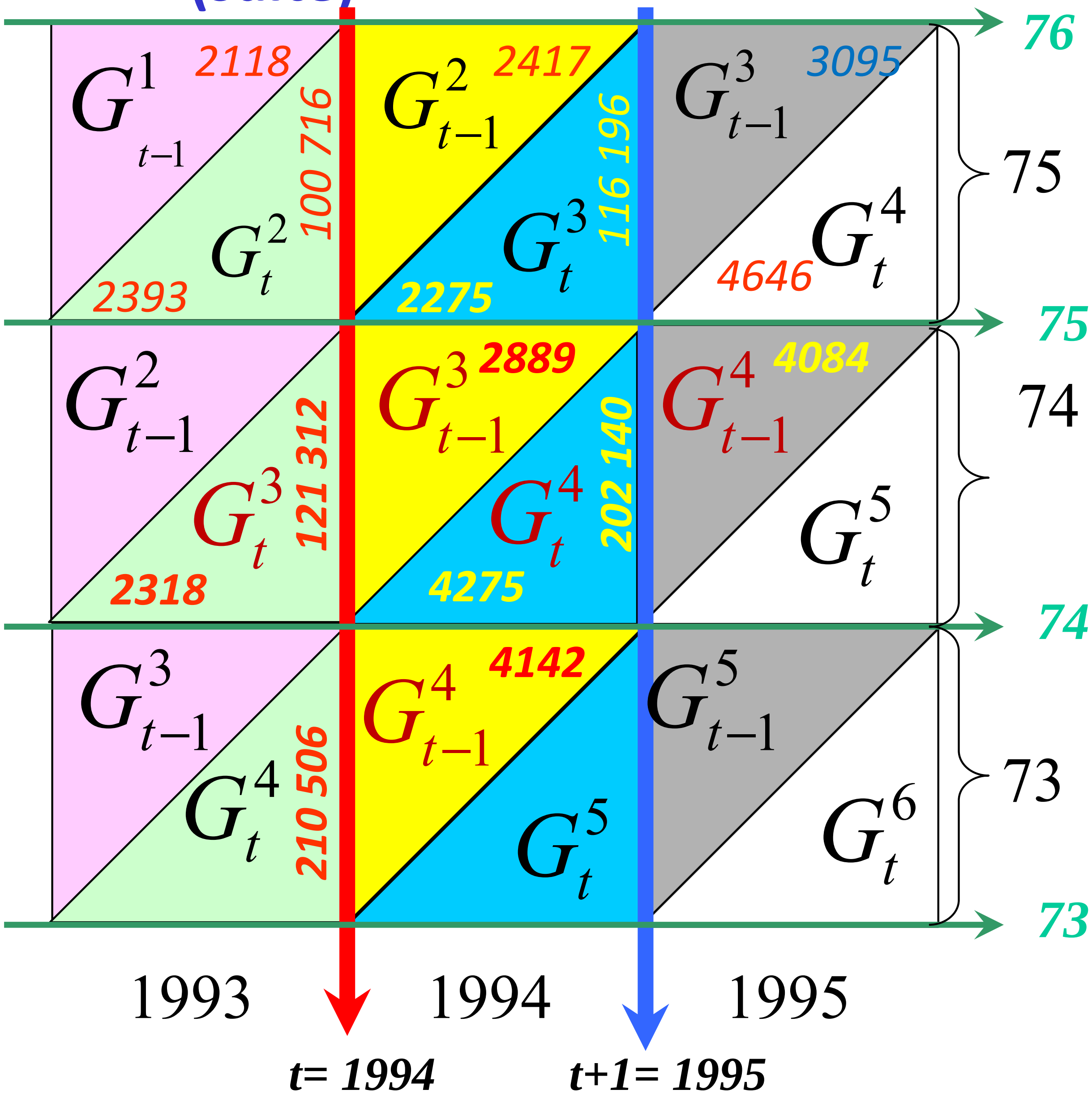
G^1 1917 = 1992-75,

G^2 1918 = 1993-75,

G^3 1919 = 1994-75,

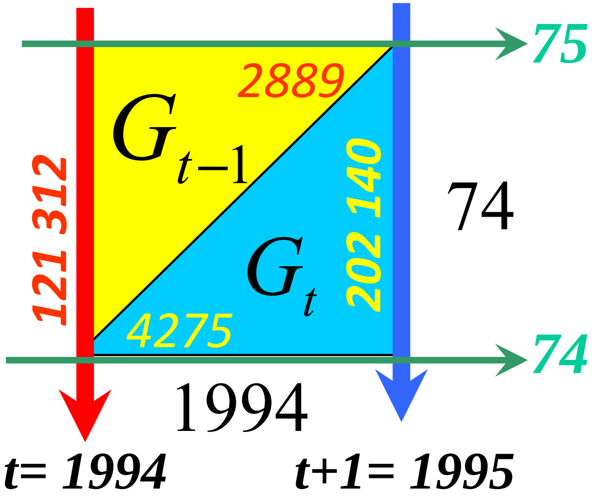
G^4 1920 = 1994-74,

G^5, G^6



Variation des effectifs de générations: conséquences pour les taux (suite).

Type I:
année. 1994
âge révolu. . . 74



G = 1994 – 74 = 1920

Dans cette G 'âge de 74 est atteint en 1994 (t)

G-1 = 1920 – 1 = 1919

Dans cette G 'âge de 74 est atteint en 1993 (t– 1)

L'axe verticale: la variation ↘, facteur 1,48
L'axe horizontale: la variation ↗, facteur 1,67

$$m^{I}_{74} = \frac{(D^{1994}_{G-1} + D^{1994}_{G})}{0,5 \cdot (P^{1994}_{74} + P^{1995}_{74})} = \frac{2 \times (2889 + 4275)}{(121312 + 202140)} = 0,04430$$

Age	1993		1994		1995		Population	
	G-1	G	G-1	G	G-1	G	1994	1995
74	2386	2318	*2889*	*4275*	4084	4280	*121312*	*202140*
75	2118	2393	2417	2275	3095	4646	100716	116196
76	2058	2163	2238	2438	2456	2376	84490	95863
77	2222	2259	2005	2137	2235	2667	74641	80091

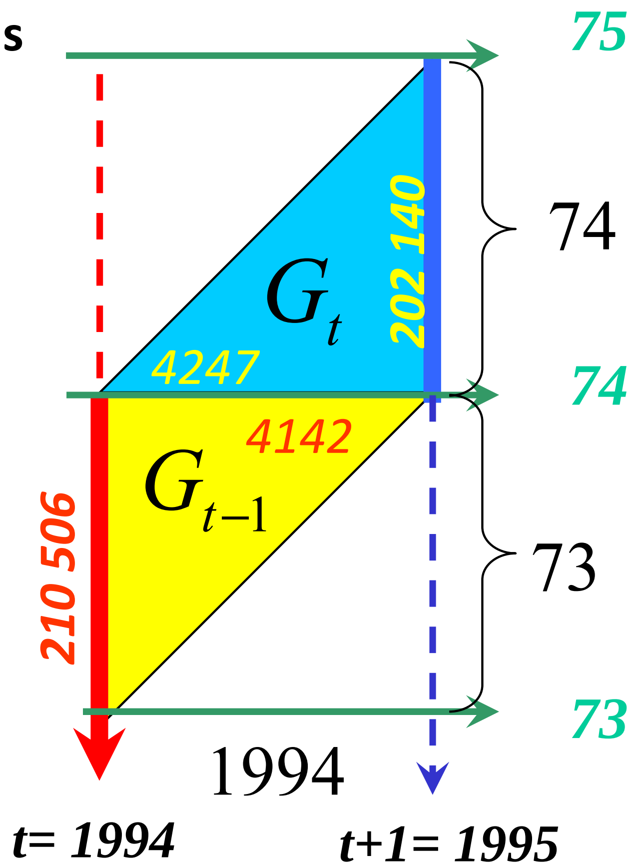
Source: « Traité » Vol. I, Ch.6. p.104

Taux type I = 44‰

Type II : **âge 74 ans en 1994 comporte deux générations par année de naissance, 1920 et 1919**

Année. 1994
Génération. . . . 1920

âge atteint en 1994 = 74



$$m_{G1920}^{II1994} = \frac{(D_{73}^{94,G-1} + D_{74}^{94,G})}{0{,}5 \cdot (P_{73}^{1994} + P_{74}^{1995})} = \frac{2 \times (4142 + 4275)}{(210506 + 202140)} = 0{,}04080$$

Age	1993		1994		Population	
	G-1	G	G-1	G	1994	1995
72	3961	3801	3648	3719	213184	206391
73	2858	4183	**4142**	3878	**210506**	205693
74	2386	2318	2889	**4275**	121312	**202140**
75	2118	2393	2417	2275	100716	116196

Taux type II′ = 41‰ (inférieur de celui de type I de ~10%)

Variation des effectifs de générations: conséquences pour les taux (suite).

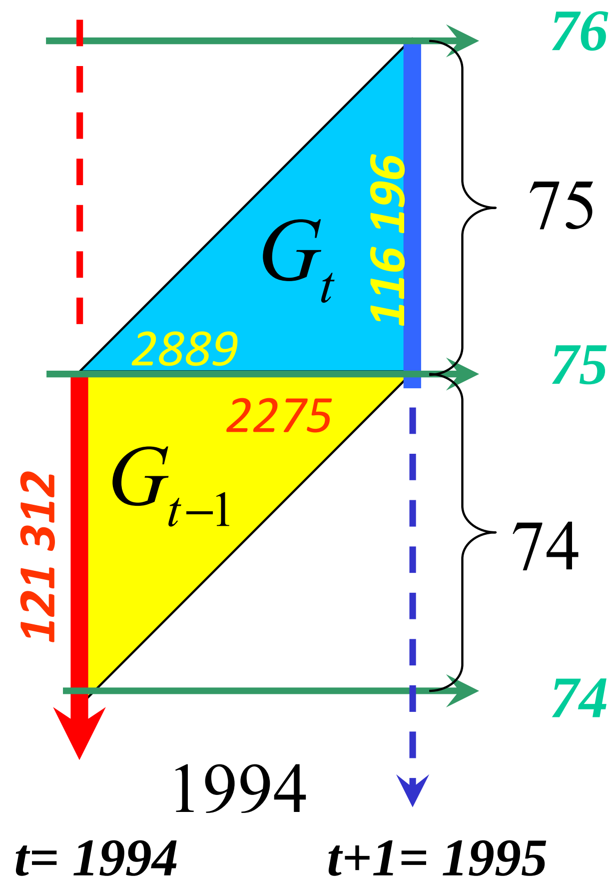
Type II:

Année………1994 ; Génération…1919

l'âge vécu en 1994 = 74 ; l'âge atteint en 1994 = 75

$$m_{G1919}^{II1994} = \frac{(D_{74}^{94,G-1} + D_{75}^{94,G})}{0{,}5 \cdot (P_{74}^{1994} + P_{75}^{1995})} =$$

$$= \frac{2 \times (2889 + 2275)}{(121312 + 116196)} = 0{,}04348$$



Notez:

121 312 – 2 275 – 2 889 = 116 148 ≠ 116 196 (effet de la migration?)

Age	1993		1994		Population début année	
	G-1	**G**	**G-1**	**G**	**1994**	**1995**
72	3961	3801	3648	3719	213184	206391
73	2858	4183	4142	3878	210506	205693
74	2386	2318	***2889***	4275	***121312***	202140
75	2118	2393	2417	***2275***	100716	***116196***

Taux type II" = 43‰

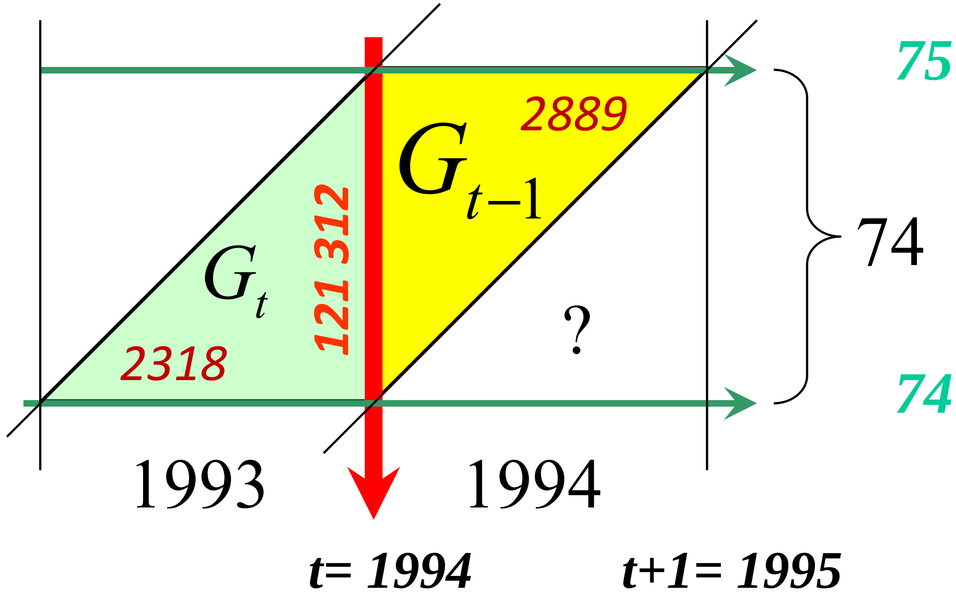
Variation des effectifs de générations: conséquences pour les taux (suite).

Type III (1e génération) :

âge 74 ans en 1994 est vécu par deux générations pour lesquelles cette âge est entendu sur deux ans de calendrier

Année. 1993-94

Génération 1919



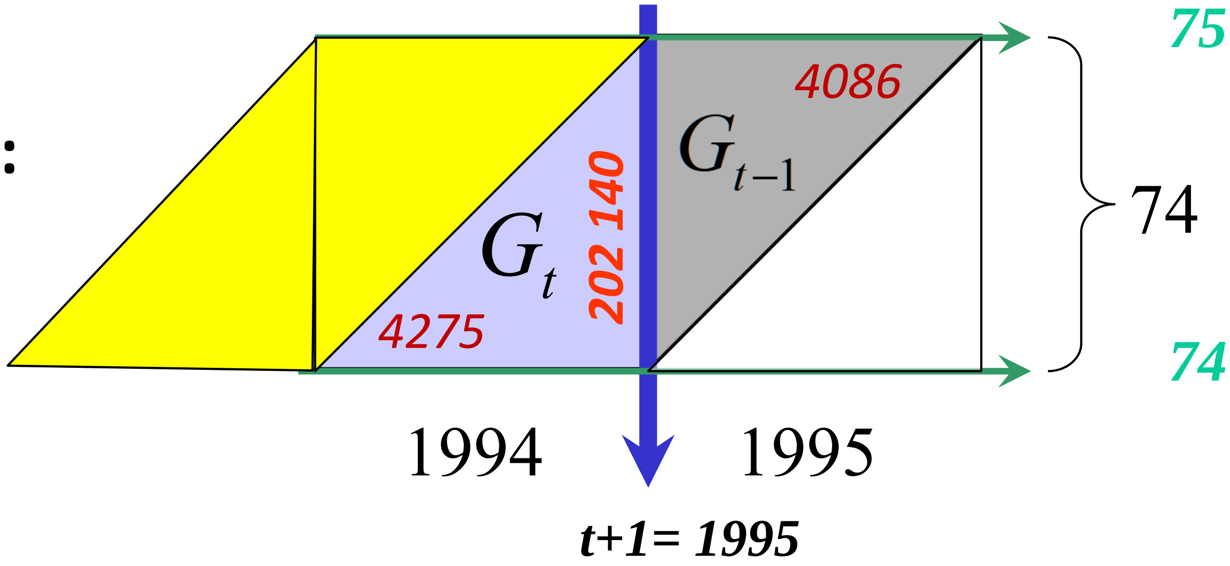
$$m_{74}^{III\ G-1919} = \frac{(D_{74}^{1993G} + D_{74}^{1994G})}{P_{74}^{1994}} = \frac{2318 + 2889}{121312} = 0{,}04292$$

Age	1993		1994		Population	
	G-1	G	G-1	G	1994	1995
72	3961	3801	3648	3719	213184	206391
73	2858	4183	4142	3878	210506	205693
74	2386	***2318***	***2889***	4275	***121312***	202140
75	2118	2393	2417	2275	100716	116196

Taux type III′ = 43‰

Variation des effectifs de générations: conséquences pour les taux (suite).

Type III (2e génération) :
Année. 1994-95
Génération…1920



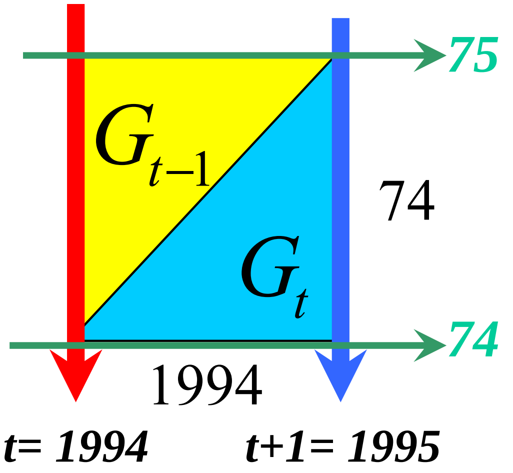
$$m_{74}^{III\ G-1920} = \frac{(D_{74}^{1994G} + D_{74}^{1995G-1})}{P_{74}^{1995}} = \frac{4275+4084}{202140} = 0{,}04135$$

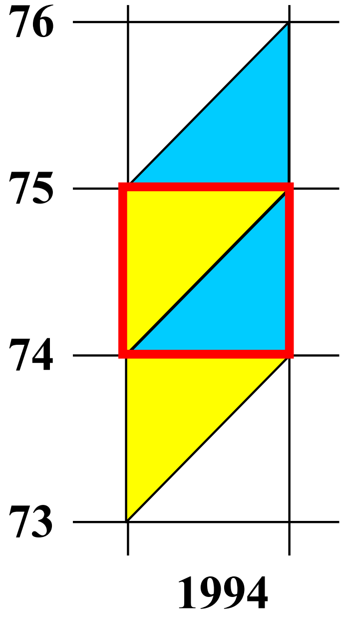
Age	1993		1994		1995		Population	
	G-1	G	G-1	G	G-1	G	1994	1995
74	2386	2318	2889	***4275***	***4084***	4280	121312	***202140***
75	2118	2393	2417	2275	3095	4646	100716	116196
76	2058	2163	2238	2438	2456	2376	84490	95863
77	2222	2259	2005	2137	2235	2667	74641	80091

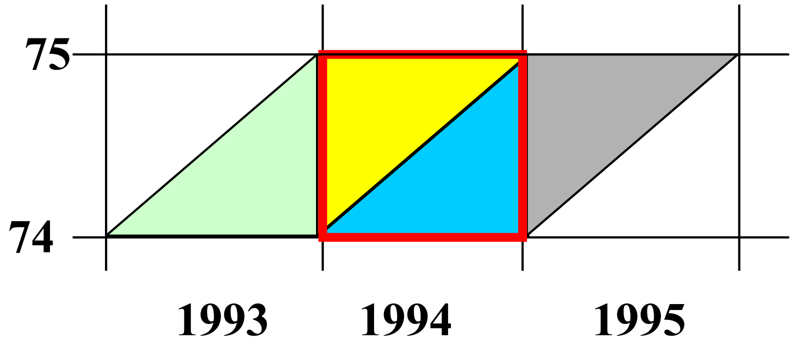
Taux type III″ = 41‰

Variation des effectifs de générations: conséquences pour les taux (suite).

Type I: année = 1994; âge = 74







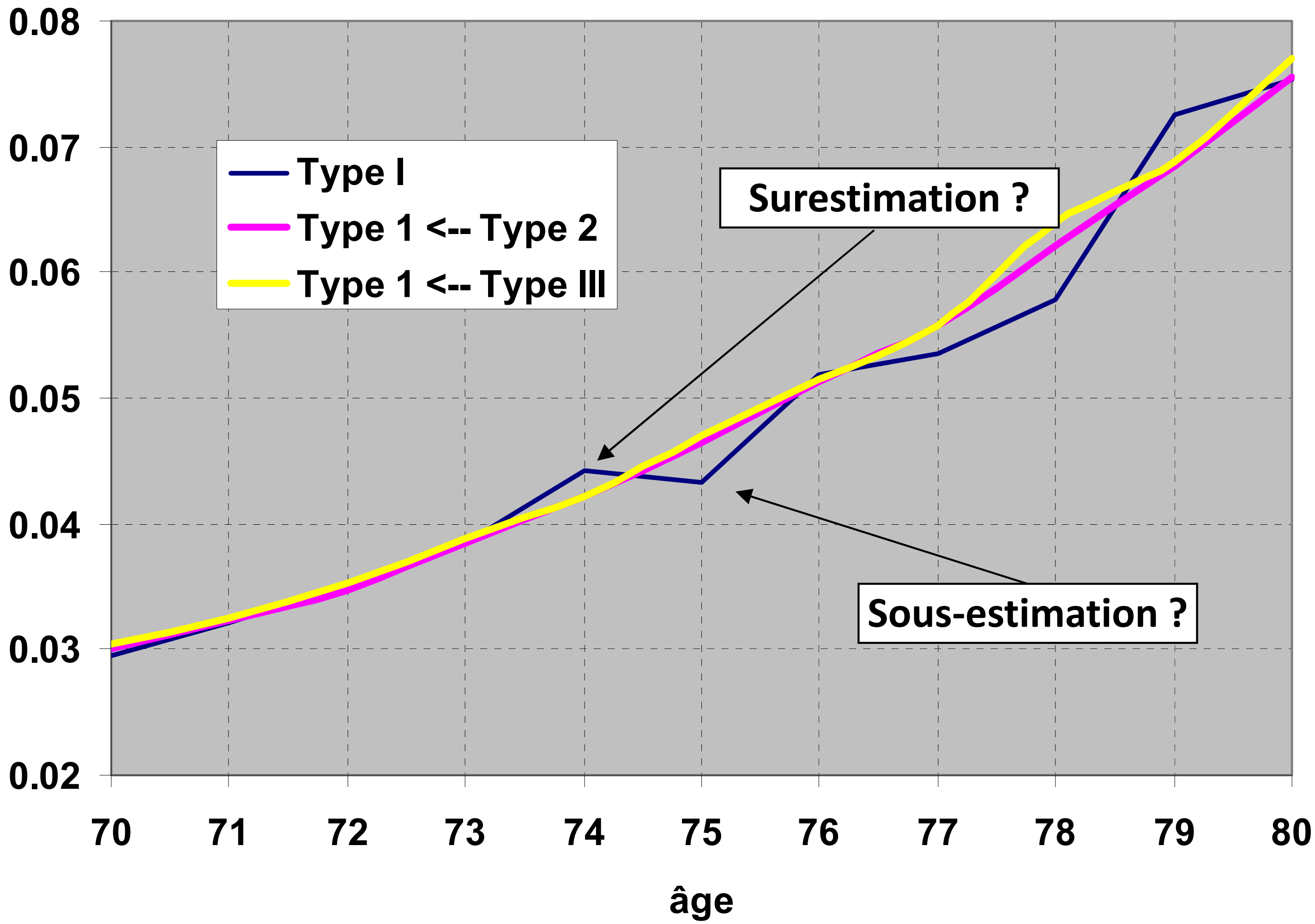
$$m_{74}^{I} = \frac{(D_{G-1}^{1994} + D_{G}^{1994})}{0{,}5 \cdot (P_{74}^{1994} + P_{74}^{1995})} = $$
$$= 0{,}04430$$

$$m_{74}^{I\,(1994)} = \frac{m_{G1920}^{II\,1994} + m_{G1919}^{II\,1994}}{2} = $$
$$= \frac{0{,}04080 + 0{,}04348}{2} = $$
$$= 0{,}04214$$

$$m_{74}^{I\,(1994)} = \frac{m_{74}^{III\,93-94} + m_{G74}^{III\,94-95}}{2} = $$
$$= \frac{0{,}04292 + 0{,}04135}{2} = $$
$$= 0{,}04214$$

Variation des effectifs de générations: conséquences pour les taux (suite)

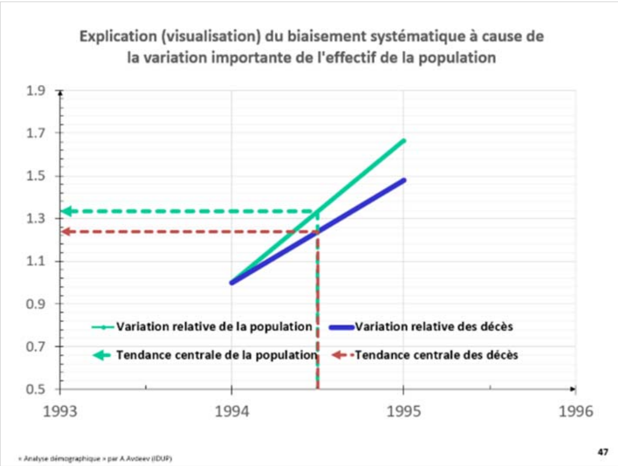
Comparaison des taux de type I calculés différemment



On observe une anomalie inexplicable sur une courbe de mortalité basée sur les taux « en carré » calculés directement à partir des données

Cette anomalie sera récurrente, si on calcule les taux de la même manière pour les années précédentes.

On peut donc faire une fausse conclusion que le risque de décès dans des certaines générations est plus fort (moins fort)

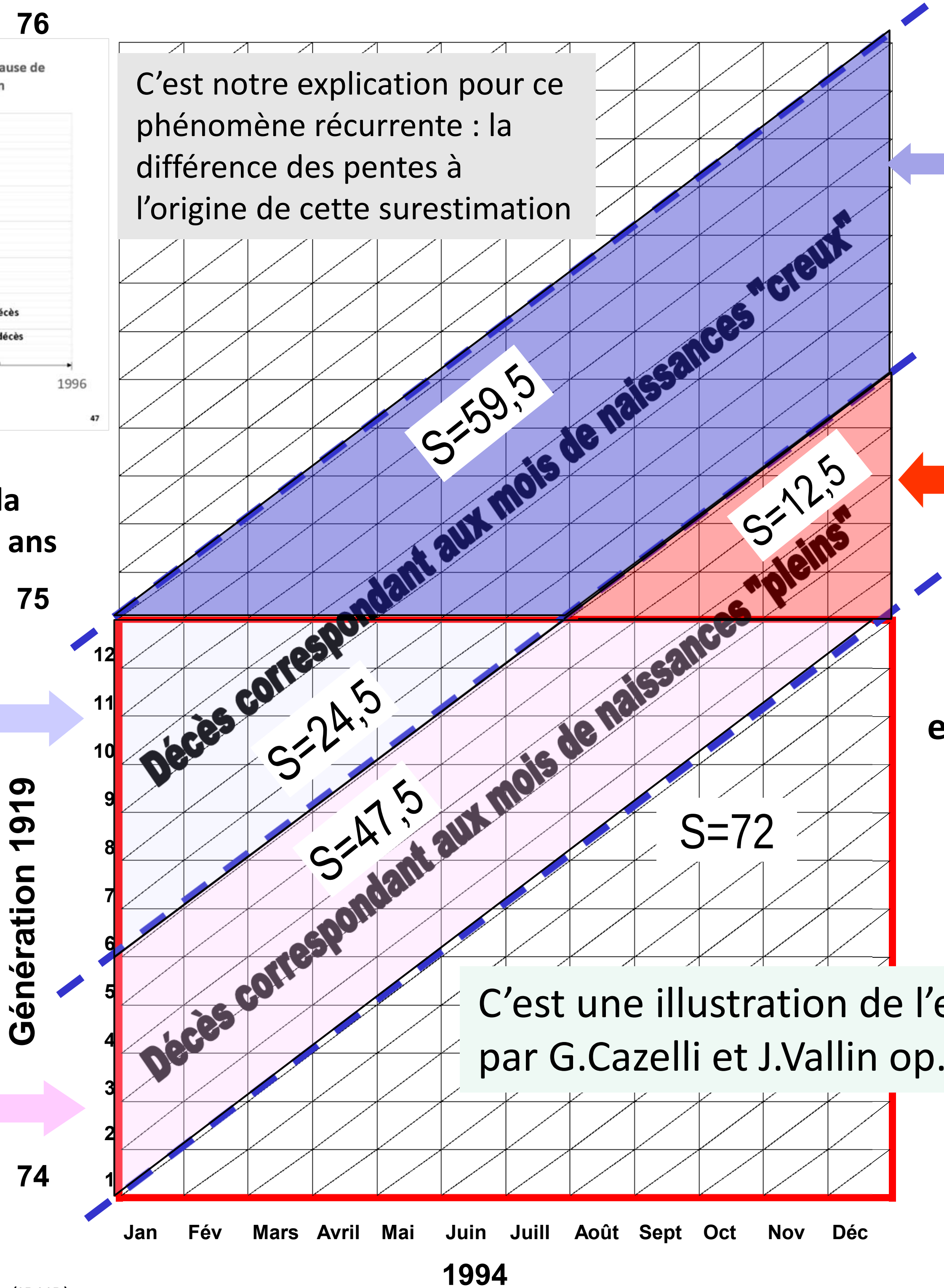


Taux de type I surestime la mortalité en 1994 à l'âge 74 ans

La portion « creuse » de la génération 1919 a vécu peu de temps entre 1 janvier 1994 et leur 75e anniversaire (S=24,5)

La portion « plaine » de la génération 1919 a vécu plus de temps entre 1 janvier 1994 et leur 75e anniversaire (S=47,5)

C'est notre explication pour ce phénomène récurrente : la différence des pentes à l'origine de cette surestimation



La portion « creuse » de la génération 1919 a vécu plus de temps en 1994 à l'âge 75 ans révolus (S=59,5), mais elle occupe plus de 50% dans la population de 74 ans au 1.1.1994

La portion « plaine » de la génération 1919 a contrubué plus aux décès l'âge 74 ans en 1994 (S=47,5), et moins à la population de 74 ans au 1.1.1994

Taux de type I sous-estime la mortalité en 1994 à l'âge 75 ans

C'est une illustration de l'explication faite par G.Cazelli et J.Vallin op.cit. p.107

Explication (visualisation) du biaisement systématique à cause de la variation importante de l'effectif de la population