

L'eterogeneità non osservata delle popolazioni*

di James W. Vaupel e Anatoli I. Yashin

3.1

L'eterogeneità in demografia

Tutte le popolazioni sono eterogenee. L'analisi demografica tradizionalmente considera due aspetti delle differenze tra individui (l'età e il sesso). Ben altre caratteristiche possono essere osservate, quali, ad esempio, la data e il luogo di nascita, il luogo (urbano o rurale) di residenza, lo stato matrimoniale, la nazionalità, la religione, il numero di figli, la dimensione della frateranza, l'età della madre o del padre alla nascita dell'individuo, la struttura della famiglia, lo stato socio-economico, il grado d'istruzione, la professione, la professione dello sposo, il consumo di tabacco, la dieta alimentare, la taglia, la qualità dell'ambiente o il luogo di residenza, lo stato di salute, le funzioni cognitive e psicologiche, il genotipo ecc. Ciononostante, anche negli studi più completi da questo punto di vista la gran parte degli attributi individuali sfugge alla misurazione, focalizzando l'attenzione su una o due delle molteplici dimensioni che differenziano tra loro gli individui.

L'eterogeneità osservata offre ai demografi numerose opportunità di analisi. L'analisi della regressione multipla, l'analisi *logit* o *probit*, le tavole di mortalità e altri metodi statistici sono stati sviluppati per stimare l'influenza delle covariate osservate. Questi metodi sono ampiamente trattati nei manuali classici di analisi demografica e sono stati largamente presentati nei capitoli precedenti.

L'eterogeneità non osservata pone problemi di analisi, mentre al contrario dovrebbe aprire la via a nuove direzioni di analisi: è una seccatura, un rompicapo, una *bestia nera*. Essa crea difficoltà

* Traduzione dall'inglese di Graziella Caselli.

ai demografi che studiano come le caratteristiche demografiche variano con l'età, con il tempo, il luogo ecc., sapendo bene che l'eterogeneità non osservata influenza lo schema dell'analisi evolutiva.

3.2

Evoluzione delle strutture

Il problema nasce dal fatto che gli individui che compongono una data coorte demografica scompaiono poco a poco, sia morendo, sia lasciando la coorte. Gli animali e le piante muoiono, le macchine si rompono, i celibi si sposano, gli sposati divorziano, le nullipare fanno un figlio, le primipare ne hanno un secondo, i figli lasciano l'abitazione dei genitori, gli studenti completano il loro ciclo di formazione, i disoccupati trovano un impiego, chi sta bene si ammala e gli ammalati guariscono. Molte analisi cercano di misurare le probabilità di transizione da uno o da un altro tipo di cambiamento di questo genere. In numerose situazioni, i demografi si preoccupano di sapere come queste probabilità di transizione si modificano con l'età: calcolano, ad esempio, probabilità di morte o di nuzialità per età. In altri casi, privilegiano la durata trascorsa da un evento precedente, considerato principale, come per la misura delle probabilità di guarire dopo una malattia o di divorziare dopo aver contratto un matrimonio. Così, nell'analisi della fecondità per rango ciò che conta è la durata trascorsa dalla nascita precedente.

Numerose analisi demografiche sono fondate sulla stima e il confronto dei tassi di uscita da coorti, la cui dimensione cambia continuamente a causa di queste uscite; il problema è che quelli che escono da una coorte hanno probabilmente una propensione ad uscire più elevata di quelli che rimangono. Le persone che muoiono ad una data età sono, indubbiamente, più fragili e maggiormente esposte ai rischi di coloro che sopravvivono fino ad età più avanzate. Le coppie che concepiscono, dopo un mese o due di tentativi, sono forse maggiormente feconde di quelle che concepiscono solo dopo numerosi mesi di tentativi. I matrimoni che finiscono rapidamente con un divorzio sono forse dei matrimoni già sbagliati in partenza. In altri termini, la composizione, la struttura eterogenea di una coorte cambia via via che la coorte riduce la sua dimensione. I più fragili o i più esposti spariscono per primi, lasciandosi dietro una coorte sempre più robusta di sopravviventi.

Tre livelli di spiegazione

L'evoluzione secondo l'età o la durata trascorsa della fecondità, della nuzialità, della mortalità o di ogni altro fenomeno di cambiamento di stato può interpretarsi da tre diversi punti di vista a seconda dei livelli di spiegazione: livello-0, livello-1 e livello-2. Il livello-0 è quello degli errori di osservazione (i dati sono errati), il livello-1 è quello dei cambiamenti prodotti a livello individuale, mentre il livello-2 è quello in cui i cambiamenti osservati corrispondono ad un artificio dovuto all'evoluzione della struttura della popolazione, cioè ad un cambiamento nel grado di eterogeneità della coorte.

Si consideri, ad esempio, il fatto, ora ben noto, che l'aumento della mortalità con l'età sta rallentando alle età molto anziane (Vaupel *et al.*, 1998; Thatcher, Kannistö, Vaupel, 1998). Una spiegazione diretta di livello-1 potrebbe essere che per individui di età elevata la probabilità di morte aumenta, in modo relativo, lentamente con l'età. Al livello-0 (dati non buoni), si potrebbe immaginare che i calcoli siano sbagliati a causa di errori nella dichiarazione delle età della popolazione, portando a sottostimare in modo crescente la mortalità all'aumentare dell'età stessa. Infine, a livello-2, il rallentamento dell'aumento della mortalità con l'età dopo 100 anni potrebbe essere sicuramente «caused by decreases in the average frailty of a population cohort at later ages as frailer members are removed by mortality»¹ (Vaupel, Manton, Stallard, 1979).

Gli esempi della decelerazione della funzione di mortalità, osservati in diverse popolazioni, sono quasi sicuramente dovuti ad una combinazione di questi tre livelli di spiegazione, le cui parti variano da un caso all'altro. In quasi tutte le popolazioni ci sono errori di dichiarazione dell'età per quelle più avanzate, e in alcuni casi questi errori sono molto importanti (Jeune, Vaupel, 1999). Tutte le popolazioni sono eterogenee, e la spiegazione di livello-2 può sempre essere richiamata in causa, anche se è difficile dire quale sia la sua portata reale. Alla fine, la spiegazione di livello-1, secondo cui gli individui invecchiano più lentamente a partire da una certa età, può essere in parte vera o completamente falsa. C'è qualche indicazione provata del fatto che a livello individuale il rischio di morte aumenta più che esponenzialmente alle età molto avanzate anche se, nella popolazione,

1. «Dovuto alla riduzione progressiva della fragilità media della coorte alle età più avanzate (alle grandi età) via via che i suoi componenti più fragili sono eliminati per morte».

il tasso di mortalità aumenta meno che esponenzialmente (Yashin, Iachine, 1997).

Almeno, dopo Edmund Halley (1693), i demografi riconoscono l'importanza dei livelli di spiegazione 0 e 2, come alternativa alla spiegazione diretta del livello-1. Ogni bravo demografo sa riconoscere la prevalenza degli errori di osservazione e tutti quelli ben addentro alle questioni sanno che i tassi demografici possono variare per il solo effetto di differenze di struttura della popolazione. Ciononostante, il livello di spiegazione 1 (ciò che si osserva a livello di popolazione vale anche a livello di individuo) sembra così naturale che anche i più meticolosi spesso si lasciano prendere, in modo un po' *ingenuo* e senza impunità, da questo tipo di interpretazione immediata delle evoluzioni o delle differenze che osservano a livello della popolazione studiata (Vaupel, Carey, 1993).

3.4 I modelli di fragilità

Si può cercare di distinguere tra le spiegazioni dei livelli 1 e 2 utilizzando una combinazione dei *modelli di fragilità* (*frailty model*, secondo Vaupel, Manton, Stallard, 1979) con i metodi statistici delle analisi di sopravvivenza (Cox, Oakes, 1984). In quest'approccio la traiettoria di un tasso di mortalità, o di uscita, può essere descritta, nell'ambito di una coorte, sia dalla funzione di sopravvivenza $s(x)$, sia dalla *funzione di rischio* $\mu(x)$, quella che i demografi chiamano "la forza di mortalità o *tasso istantaneo di mortalità*" (qualche volta si parla anche, in alcuni contesti, di *funzione di intensità*). La funzione di sopravvivenza e la funzione di rischio sono legate tra loro mediante le due seguenti formule:

$$[1a] \quad \mu(x) = - \frac{ds(x)/dx}{s(x)}$$

e

$$[1b] \quad s(x) = e^{-\int_0^x \mu(t) dt}$$

Nel caso più semplice, non si conosce nessun'altra caratteristica individuale se non l'età (indipendentemente dalle sue caratteristiche, la coorte è considerata come un tutt'uno, così come "l'insieme degli uomini nati in Francia nel 1948").

Poiché tutte le popolazioni sono eterogenee, si può utilmente considerare la popolazione come combinazione di sotto-popolazioni omogenee (che possono, al limite, ridursi ognuna ad un solo individuo). Si chiami $s(x,z)$ la funzione di sopravvivenza della sotto-popolazione con "fragilità" z , in cui la fragilità, in questo caso, è semplicemente la sua suscettibilità all'esposizione al rischio in questione. Generalmente, i modelli di fragilità sono costruiti in modo tale che più l'individuo è fragile, più è suscettibile o esposto al rischio in causa.

Sia $\bar{s}(x)$ la funzione di sopravvivenza della popolazione totale, tale che, nel continuo sia:

$$[2a] \quad \bar{s}(x) = \int_0^{\infty} s(x,z)g(z)dz$$

in cui $g(z)$ è la funzione di distribuzione delle probabilità di z all'età zero, e, nel discreto:

$$[2b] \quad \bar{s}(x) = \sum_z \pi(z)s(x,z)$$

in cui $\pi(z)$ è la proporzione della coorte appartenente alla sotto-popolazione z all'età zero. Secondo i casi, questo modello generale di fragilità può essere formulato in diversi modi.

3.4.1. I modelli di rischio relativo

Una formulazione possibile per il *modello di rischi proporzionali*, chiamato anche *modello di rischio relativo*, è:

$$[3a] \quad \mu(x,z) = z\mu(x)$$

che, come hanno suggerito James Vaupel, Kenneth Manton ed Eric Stallard (1979), può essere scritta anche:

$$[3b] \quad s(x,z) = s(x)^z$$

In questo modello, $\mu(x)$ è la funzione del rischio di base, o funzione *standard*, il rischio per gli individui di fragilità 1, mentre $s(x)$ è la fun-

zione di sopravvivenza corrispondente. In effetti, Vaupel, Manton e Stallard hanno mostrato che:

$$[4] \quad \bar{\mu}(x) = \bar{z}(x)\mu(x)$$

in cui $\bar{z}(x)$ è la fragilità media di quelli che sopravvivono all'età x . Poiché z è dato e non varia con l'età, $\bar{z}(x)$ diminuisce con l'età via via che i più fragili escono dalla coorte. Perciò, $\bar{\mu}(x)$ aumenta più lentamente di $\mu(x)$. In effetti, $\bar{\mu}(x)$ può diminuire anche se $\mu(x)$ aumenta.

In questo modello, z è spesso ipotizzato seguire una legge gamma di media τ e varianza σ^2 , poiché una tale distribuzione conduce a relazioni matematiche molto pratiche. In particolare, se la fragilità segue una legge gamma, si ha:

$$[5] \quad \bar{z}(x) = (\tau + \sigma^2 \int_0^x \mu(t) dt)^{-1} = \bar{s}(x)^{\sigma^2}$$

in cui $\bar{s}(x)$ è la funzione di sopravvivenza media della popolazione totale. Dalla [5] ne consegue, quindi, che:

$$[6] \quad \bar{s}(x) = (\tau - \sigma^2 \int_0^x \mu(t) dt)^{-1/\sigma^2}$$

A titolo di esempio del modello di rischi relativi, con fragilità che segue una legge gamma, si ipotizzi che la mortalità segua, a livello individuale, una funzione di Gompertz:

$$[7] \quad \mu(x) = ae^{bx}$$

In questo caso, dalle formule [4] e [5] deriva che la curva (o traiettoria) di mortalità della popolazione segue un modello logistico:

$$[8] \quad \bar{\mu}(x) = \frac{ae^{bx}}{1 + \frac{a\sigma^2}{b}(e^{bx} - 1)}$$

la cui crescita con l'età rallenta al di là del valore b/σ^2 .

3.4.2. I modelli d'invecchiamento accelerato

Un altro caso particolare di modello di fragilità è il *modello d'invecchiamento accelerato*, in cui:

$$[9] \quad \mu(x,z) = \mu(xz)$$

che è della stessa natura del modello di fragilità accelerata utilizzato in ingegneria per lo studio dell'affidabilità. Nel caso particolare in cui $\mu(x)$ segua una traiettoria di Weibull della forma:

$$[10] \quad \mu(x) = ax^b$$

in cui a e b sono dei parametri, questo modello equivale al modello di rischio relativo

$$[11] \quad a(zx)^b = z^b ax^b = z^b \mu(x)$$

Nel caso particolare in cui $\mu(x)$ segua una traiettoria gompertziana, come quella della formula [7], il modello di invecchiamento accelerato è della forma ae^{zbx} , mentre il modello di rischio relativo assume la forma zae^{bx} . Piccoli cambiamenti nell'inclinazione di b possono produrre effetti sulla mortalità alle grandi età maggiori di quelli di grandi cambiamenti nel livello di a . Ne consegue che, per produrre sostanziali differenze tra $\mu(x)$ e $\bar{\mu}(x)$ alle età più avanzate (FIG. 3.1) è richiesta un'eterogeneità molto minore nei modelli gompertziani di invecchiamento accelerato che in un modello gompertziano di rischio relativo.

3.4.3. I modelli di fragilità discontinua

L'utilizzo del modello di fragilità discontinua rappresenta un modo molto pratico per giungere all'approccio generale della fragilità (Vaupel, Yashin, 1985). In questo caso si ha che:

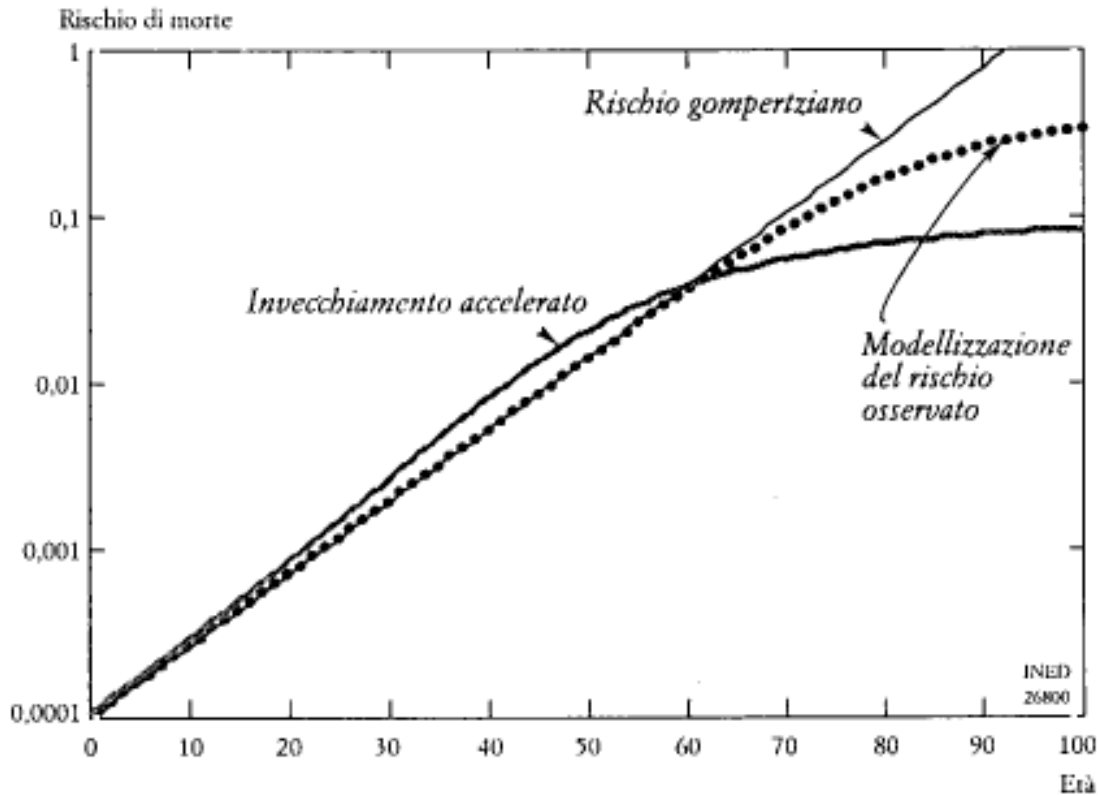
$$[12] \quad \mu(x,z) = \mu_z(x)$$

In questa espressione, z è un indicatore di fragilità per diverse sottopopolazioni, ognuna con una propria funzione di rischio. Se si chiama π_z la proporzione della popolazione appartenente alla sottopopolazione z all'età zero, si ha:

$$[13] \quad \bar{s}(x) = \sum_z \pi_z s_z(x)$$

FIGURA 3.1

Il modello di invecchiamento accelerato può produrre una decelerazione più forte nella traiettoria di mortalità con meno eterogeneità di quella richiesta per il modello di rischio relativo



Basi: rischi gompertziani con $a = 0,0001$ e $b = .1$ confrontati alla modellizzazione dei rischi osservati nella popolazione partendo da un modello di rischio relativo con $s_2 = 0,25$ e con un modello di invecchiamento accelerato con $s_2 = 0,05$. N.B.: le curve sono rappresentate su scala logaritmica.

e

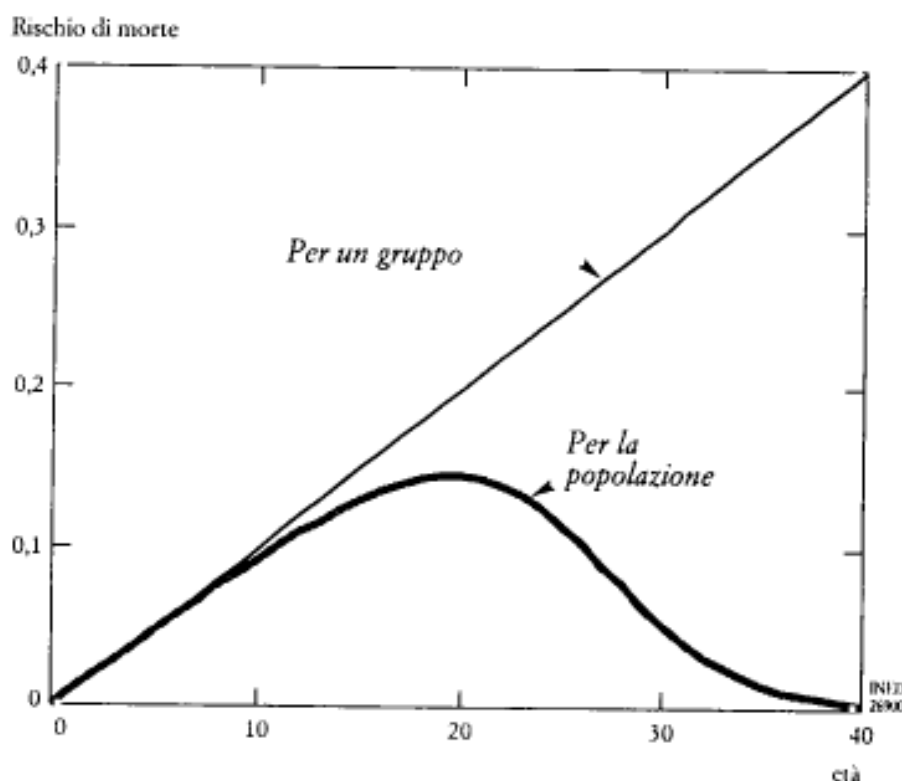
$$[14] \quad \bar{\mu}(x) = \frac{\sum_z \pi_z s_z(x) \mu_z(x)}{\sum_z \pi_z s_z(x)}$$

Se si ipotizza che z sia un fattore di rischio relativo, deriva che $\mu_z(x) = z\mu(x)$ e $s_z(x) = s(x)^z$. James Heckman e Burton Singer (1984) hanno suggerito di utilizzare questo modello per controllare gli effetti dell'eterogeneità non osservata al momento dell'adattamento dei modelli ai dati. Più generalmente, tuttavia, $\mu_z(x)$ può prendere una forma diversa per ogni valore di z .

Un esempio molto semplice di modello di fragilità discontinua è il modello "mover-stayer" (cambiamento-continuità) di Blumen *et al.* (1955), nel quale, nell'ambito di una popolazione, un gruppo è su-

FIGURA 3.2

Il rischio per la popolazione (curva in grassetto) può aumentare e poi diminuire se il rischio per un gruppo (tratto fino, con $m(x) = .01x$) aumenta e l'altro gruppo è immune



Si è ipotizzato che il 95% della popolazione sia fragile.

scettibile di emigrare, di sposarsi, di divorziare o di contrarre questa o quella malattia, mentre l'altro gruppo è immune. Se π è la proporzione di popolazione fragile, si può dedurre dalla [14] che:

$$[15] \quad \bar{\mu}(x) = \pi s(x) \mu(x) / (\pi s(x) + 1 - \pi)$$

Anche se $\mu(x)$ aumenta rapidamente, $\bar{\mu}(x)$ potrà diminuire quando $s(x)$ si approssima a zero. Ciò è illustrato nella FIG. 3.2.

I tassi di divorzio, in alcuni paesi, in certi periodi, seguono lo schema ascendente-discendente illustrato nella FIG. 3.2. Questo implica che i matrimoni nei primi anni sono meno stabili? Non necessariamente, come è noto. Lo stesso risultato potrebbe prodursi se il secondo gruppo non fosse totalmente immune ma semplicemente sottoposto a un rischio minore. In effetti, uno schema ascendente-discendente dello stesso tipo potrebbe essere ottenuto con un rischio forte-

mente crescente per il gruppo più esposto e fortemente decrescente per il gruppo a minor rischio. Per un gruppo, i matrimoni si consolidano con il tempo, mentre per l'altro, diventano più vulnerabili (a scapito delle apparenze illustrate dalla curva media della coorte che non evidenzia picchi di divorzio).

3.4.4. I modelli con fragilità variabile

Come ha scritto George Box (1976), tutti i modelli sono falsi, ma alcuni sono utili. È spesso utile ipotizzare costante la fragilità di un individuo, almeno partendo da una certa età, e classificare gli individui in gruppi secondo la loro fragilità a quell'età. Alternativamente, può anche essere utile sviluppare modelli nei quali la fragilità di un individuo varia con il tempo o l'età.

Nel caso più semplice, per un modello di questo tipo, tutti gli individui partono con fragilità uguale ad uno. Essi subiscono un rischio di morte $\mu_1(x)$ all'età x . Sono, quindi, sottoposti a un rischio $\lambda(x)$ che la loro fragilità passi da 1 a 2, nel qual caso il loro rischio di morte passerà a $\mu_2(x)$. Il secondo stato potrebbe essere associato a qualche evento morboso, come un attacco cardiaco o la perdita della capacità di camminare. Si potrebbe immaginare la stessa situazione a proposito del rischio di divorzio all'arrivo di un figlio. Sia $s_1(x)$ la proporzione della coorte che sopravvive, con fragilità 1 all'età x e, in modo simile sia $s_2(x)$ la proporzione della coorte che sopravvive con fragilità 2 all'età x ; nel caso più semplice in cui le tre funzioni di rischio siano costanti, si può facilmente mostrare che:

$$[16a] \quad s_1(x) = e^{-(\mu_1 + \lambda)x}$$

e che:

$$[16b] \quad s_2(x) = \frac{\lambda}{\mu_1 + \lambda - \mu_2} (e^{-\mu_2 x} - e^{-(\mu_1 + \lambda)x})$$

Il rischio per la popolazione è, quindi, dato da:

$$[17] \quad \bar{\mu}(x) = \frac{\lambda \mu_1 e^{(\mu_1 + \lambda - \mu_2)x} + (\mu_1 - \mu_2)(\mu_1 + \lambda)}{\lambda e^{(\mu_1 + \lambda - \mu_2)x} + (\mu_1 - \mu_2)}$$

All'età zero, $\bar{\mu}(0) = \mu_1$ e quando x tende all'infinito, $\bar{\mu}(x) = \mu_2$ se $\mu_2 \geq \mu_1 + \lambda$ e $\bar{\mu}(x) = \mu_1 + \lambda$ negli altri casi.

Hervé Le Bras (1976), come Leonid Gavrilov e Nathalia Gavrilova (1991), hanno proposto delle generalizzazioni di questo modello. Al posto di due stati di fragilità, si può ipotizzare che la fragilità z possa prendere qualsiasi valore intero non negativo. Alla partenza, tutti hanno una fragilità uguale a zero. Le persone con fragilità z subiscono un rischio di morte di $\mu_0 + z\mu$, e corrono un rischio $\lambda_0 + z\lambda$ che la loro fragilità passi a $z+1$. Anche se μ_0 , μ , λ_0 e λ sono delle costanti e non variano né con l'età né con il tempo, il rischio $\bar{\mu}(x)$ per la popolazione segue una curva logistica. Questa curva è identica a quella ottenuta quando la fragilità segue una legge gamma fissa e il rischio di base è di forma gompertziana $\mu(x) = ac^{bx}$, con $\mu(x,z) = z\mu(x) + c$, in cui c è il termine costante di Makeham (Yashin, Vaupel, Iachine, 1994). Senza informazioni ulteriori è impossibile dire se la fragilità sia fissa o variabile.

Anziché attenersi a valori discreti, si può anche modellizzare la fragilità in modo tale che vari continuamente. Si può, ad esempio, costruire un modello di fragilità che si modifichi sulla base di un'equazione differenziale stocastica per chiarire l'interazione tra indebolimento, recupero, selezione e invecchiamento (Vaupel, Yashin, Manton, 1988). Il modello richiede grande intuizione per spiegare le conseguenze sulla mortalità di disastri come le guerre, la fame o le epidemie, che riducono la resistenza dei sopravvissuti. Un risultato chiave è che l'indebolimento (o fragilizzazione) e la selezione sono interdipendenti: l'indebolimento che accresce l'eterogeneità della popolazione porterà ad una selezione per morte; la selezione, che altera la distribuzione della fragilità, influenzerà l'effetto degli eventi che producono l'indebolimento. L'equazione base di questo modello è la seguente:

$$[18] \quad \mu(x,z) = \mu_0(x) + z(x)\mu^*(x)$$

in cui $\mu_0(x)$ è il rischio di base, mentre $\mu^*(x)$ determina il rischio addizionale e $z(x)$ è la fragilità dell'individuo di età x tale che:

$$[19] \quad z(x) = Y^2(x)$$

in cui $Y(0)$ segue la legge normale e:

$$[20] \quad dY(x) = [a_0(x) + (a_1(x) - a_2(x))Y(x)]dx + b(x)dW(x)$$

in cui W è un processo di Wiener con $W(0) = 0$. Le funzioni a_0 e a_1 rappresentano gli effetti d'indebolimento mentre a_2 rappresenta il ri-

torno allo stato precedente, il recupero; la funzione b determina l'importanza del termine del processo di Wiener.

3.4.5. I modelli di fragilità correlata

Condividendo gli stessi geni e lo stesso ambiente durante gli anni della prima infanzia, due gemelli possono avere la stessa fragilità. Più generalmente, parenti o persone che vivono nello stesso ambiente possono avere fragilità simili. Modelli a fragilità condivisa (Vaupel, 1991a e 1991b) possono essere utilizzati per analizzare situazioni di questo tipo, ma i *modelli di fragilità correlata* sono i più appropriati e i più potenti (Yashin, Vaupel, Iachine, 1995; Yashin, Iachine, 1997).

Una forma semplice di questi modelli fa riferimento al caso particolare di coppie di gemelli, in cui uno dei due ha una data fragilità z_1 , mentre l'altro ha una fragilità z_2 e il cui rischio di morte è dato da:

$$\mu(x, z_i) = z_i \mu_0(x), \quad i = 1, 2$$

La correlazione tra le due fragilità viene modellizzata come di seguito.

Si ponga:

$$[21a] \quad z_1 = y_0 + y_1$$

e:

$$[21b] \quad z_2 = y_0 + y_2$$

in cui le y_i , $i = 0, 1, 2$ sono tre variabili aleatorie indipendenti, che seguono leggi gamma aventi lo stesso parametro di scala. Le distribuzioni gamma di y_1 e y_2 hanno lo stesso parametro di forma, ma questo parametro può differire per y_0 . Le fragilità z_1 e z_2 sono per definizione di media 1 e della stessa varianza σ^2 . I valori di σ^2 e ρ , la varianza e il coefficiente di correlazione tra le due fragilità sono semplici funzioni dei parametri di scala della forma.

La funzione di sopravvivenza bivariata per la popolazione dei gemelli è data da (Yashin, Vaupel, Iachine, 1995):

$$[22] \quad \bar{s}(x_1, x_2) = s(x_1)^{1-\rho} \bar{s}(x_2)^{1-\rho} (\bar{s}(x_1)^{-\sigma^2} + \bar{s}(x_2)^{-\sigma^2} - 1)^{-\rho/\sigma^2}$$

Poiché la sopravvivenza dei gemelli adulti è molto simile a quella degli adulti non gemelli, per studiare la mortalità degli adulti si può considerare la funzione $\bar{s}(x)$ tratta dalle tavole di mortalità dell'insie-

me della popolazione. In questo caso, la funzione di sopravvivenza bivariata non dipende che da σ^2 e da ρ , e non è necessaria alcuna ipotesi sulla forma della funzione del rischio di base. Utilizzando la [22] i valori di σ^2 e di ρ possono essere stimati con la procedura di massima verosimiglianza descritta di seguito. Inoltre, la funzione di sopravvivenza base può essere calcolata riadattando la [6] come segue:

$$[23] \quad s(x) = \exp((1 - \bar{s}(x)^{-\sigma^2}) / \sigma^2)$$

e $\mu(x)$ può essere calcolato partendo dalla [1]. Applicando questo modello ai dati disponibili sulla sopravvivenza dei gemelli danesi nati tra il 1870 e il 1900, si è potuto dimostrare (Yashin, Iachine, 1997) che il rischio di morte di base aumenta più che esponenzialmente con l'età oltre i 30 anni, anche se per la popolazione intera si osserva una decelerazione alle età più elevate. In altre parole, può essere che questa decelerazione dipenda interamente da una spiegazione di livello-2 (cambiamento di struttura dovuto alla selezione per morte) e che, a livello individuale, la traiettoria reale della mortalità aumenti più velocemente di una curva di Gompertz.

3.5 Dati empirici

I dati sulla sopravvivenza o sulla durata in uno stato utilizzati nei modelli di fragilità sono generalmente della forma seguente: si ha una coorte di n individui di cui si conoscono le età al decesso $X_i, i = 1, \dots, n$. Più generalmente, X_i è l'età ad un evento qualsiasi, un matrimonio ad esempio, o anche la durata trascorsa tra un dato evento, come il matrimonio, e un altro, come il divorzio. Per comodità, in questo esempio si considera X_i come un'età alla morte.

L'età alla morte può non essere nota per tutti gli individui: si può, ad esempio, conoscere solamente se un determinato individuo era ancora in vita ad una data età. Si parla in questo caso di informazione censurata (o troncata) a destra. Questo può succedere se alcuni individui non "muoiono" mai (o, ad esempio, se alcune donne non fanno figli, alcune persone non si sposano mai, alcuni individui non divorziano mai ecc.). Questo può succedere se alcuni individui sono ancora in vita alla fine del periodo di osservazione, o se alcuni individui escono dal campo d'osservazione prima di morire.

Può anche accadere che si sappia solamente che l'individuo i , è deceduto tra l'età x_1 , e l'età x_2 . E, può anche succedere che un indi-

viduo non sia seguito dalla nascita, ma solamente dopo un'età x_0 , in modo tale che l'individuo non è sottoposto al rischio di morte che dopo questa età x_0 . Si dice, allora, che i dati sono troncati – o censurati – a sinistra.

Possono essere osservate numerose covariate: si può chiamare w_{ij} il valore della j -ima covariata per un individuo i .

3.6

Metodi di stima dei parametri

Per l'adattamento dei modelli di fragilità ai dati empirici di sopravvivenza o di durata trascorsa dopo un evento anteriore, si possono utilizzare diversi metodi di stima. In un importante articolo su *Deceleration in the Age Pattern of Mortality at Older Ages*², Shiro Horiuchi e John Wilmoth (1998) hanno stimato i parametri di un modello Gompertz-Makeham con effetti di periodo ed una fragilità non osservata con una procedura dei minimi quadrati ponderati. Essi hanno criticato puntigliosamente il metodo utilizzato, tanto che il loro articolo presenta nello stesso tempo un grande valore pedagogico e un alto interesse scientifico.

Generalmente, però, nell'analisi dei dati di sopravvivenza o di durata in uno stato e in particolare nell'adattamento dei modelli di fragilità, si utilizza il metodo della massima verosimiglianza. Per la discussione di quest'approccio, la cosa migliore è raccomandare di far ricorso ad un buon manuale, come, ad esempio, quello di David Cox e D. Oakes (1984). In effetti, in questa sede ci si limita ad affrontare alcuni aspetti particolarmente importanti della stima dei parametri nei modelli di fragilità.

La verosimiglianza di un'osservazione X_i può essere assimilata alla probabilità di osservare questo valore nell'ambito di un modello particolare con valori parametrici dati. Più generalmente, la verosimiglianza può essere proporzionale anziché essere uguale alla probabilità, poiché qualsiasi valore parametrico massimizzante la probabilità massimizzerà anche ogni quantità che le sia proporzionale. Sia $s_i(x)$ la probabilità di sopravvivere tra l'età zero e l'età x , per un individuo con un vettore di covariate w_i e una fragilità non osservata z , se l'età al decesso osservata è X_i e se l'individuo è sopravvissuto dall'età zero, la verosimiglianza dell'età alla morte osservata è $\mu_i(X_i)s_i(X_i)$. Se si sa solamente che l'individuo è o meno sopravvissuto fino all'età X_i , allo-

2. *Decelerazione della curva di mortalità alle età avanzate.*

ra la verosimiglianza di questa osservazione sarà $s_i(X_i)$. Se si sa che l'individuo è deceduto tra le età X_{1i} e X_{2i} , allora la verosimiglianza è $s(X_{1i}) - s(X_{2i})$. Se, infine, l'individuo è stato osservato per la prima volta all'età X_{1i} ed poi è morto all'età X_{2i} , la verosimiglianza sarà $\mu(X_{2i})s(X_{2i}) / s(X_{1i})$.

Nell'analisi di sopravvivenza è normale considerare il logaritmo della verosimiglianza, il *log-verosimiglianza*, poiché la verosimiglianza di una serie di dati è spesso estremamente piccola. Se $L(X_i)$ è il log-verosimiglianza dell'osservazione, il log-verosimiglianza dell'intera serie di dati è fornito dalla somma di $L(X_i)$. La stima del valore di un parametro di un modello mediante la massima verosimiglianza è la stima che massimizza la verosimiglianza ed è anche il log-verosimiglianza dei dati.

L'effetto sulla sopravvivenza delle covariate osservate può essere modellato in numerosi modi. Poiché in questo caso ci si riferisce all'eterogeneità non osservata, e non ai metodi classici delle analisi di sopravvivenza, si limiterà l'attenzione al caso semplice in cui le covariate sono fisse (piuttosto che mutevoli con il tempo). Inoltre, si ipotizzerà che l'effetto delle covariate sul rischio individuale sia dato da $W_i \mu_i(x, z)$, in cui W_i è il rischio relativo netto imposto dal vettore delle covariate osservate. Spesso nell'analisi di sopravvivenza, W_i è modellato da

$$W_i = e^{\sum_j b_j w_j}$$

in cui le b_j sono dei coefficienti stimati.

Per comodità e per una maggiore chiarezza si considera solamente il rischio relativo di un modello di fragilità, con una funzione di fragilità che segue una legge gamma. Per il modello di rischio relativo (cfr. formula [3b]),

$$[24] \quad s_i(x) = s(x)^{w_i \alpha}$$

in cui $s(x)$ è la funzione standard di sopravvivenza per individui di fragilità 1 con un rischio relativo stimato W_i di 1. Nel caso particolare di una fragilità che segue una legge gamma, secondo la formula [6], si ha:

$$[25] \quad \bar{s}_i(x) = (1 - \sigma^2 W_i \ln s(x))^{-1/\sigma^2}$$

in cui $\bar{s}_i(x)$ è la probabilità per un individuo con rischio relativo sti-

mato W_i , di sopravvivere all'età x e con varianza della fragilità σ^2 . La barra sulla s indica che \bar{s}_i è una media: la fragilità non osservata z è ricavata dalla formula prendendo il valore atteso di s in funzione di z . Il valore corrispondente di $\bar{\mu}^i(x)$ può essere calcolato grazie alla formula [1] e il log-verosimiglianza $L(x)$ può, quindi, essere ottenuto come indicato precedentemente.

Ora si deve sapere come stimare la funzione standard di sopravvivenza $s(x)$. Per questo ci sono due vie possibili. La prima è di ipotizzare una forma parametrica. Ad esempio, si può supporre che $s(x)$ (e $\mu(x)$) abbiano una forma gompertziana o weibulliana (per un esempio di questo tipo di analisi, cfr. Manton, Stallard, Vaupel, 1986).

L'altra soluzione è quella di stimare $s(x)$ in modo non parametrico, partendo da una serie di età che copra un certo intervallo senza fare ipotesi sulla forma della curva delle $s(x)$. Recentemente sono stati proposti molti metodi e la ricerca in questo campo avanza a grandi passi. Per ora ci si può accontentare di presentare uno di questi metodi per illustrare l'idea generale sulla quale posa la stima non parametrica.

Si ipotizzi che i dati disponibili per l'analisi siano quelli di una grande indagine campionaria rappresentativa della popolazione e che, dall'altra parte, sia nota la mortalità della popolazione attraverso le statistiche di stato civile. Sia $\bar{s}(x)$ la curva di sopravvivenza della popolazione e, come già visto, $s_i(x)$ la funzione di sopravvivenza per un individuo i dell'indagine. Per un campione casuale molto ampio, si può senza tema di commettere un grande errore, considerare la seguente equazione:

$$[26] \quad \bar{s}(x) = \sum_{i=1}^n s_i(x) / n$$

Si può quindi applicare il metodo seguente, cioè per un rischio relativo con una fragilità distribuita secondo una funzione gamma, la formula [25] può essere sostituita dalla [26], che porta a:

$$[27] \quad \bar{s}(x) = \sum_i (1 - \sigma^2 W_i \ln s(x))^{-1/\sigma^2} / n$$

Per ogni serie di valori stimati di W_i e σ^2 , uno ed uno solo dei valori di $s(x)$ soddisfa l'equazione. Così, σ^2 e i coefficienti che determinano W_i possono essere determinati con il metodo di massima verosimiglianza sotto i vincoli imposti nella [27].

Le proprietà teoriche e pratiche di questo algoritmo devono ancora essere oggetto di investigazione. Si stanno sviluppando numerose altre procedure di stima che non richiedono stime parametriche di $s(x)$ e potrebbero anche essere utilizzati diversi metodi di imputazione, altri algoritmi EM e altri nuovi concetti (cfr., ad esempio, Andersen *et al.*, 1993). La stima dei modelli di fragilità è attualmente un punto cruciale nella ricerca statistica ed il prossimo decennio conoscerà senza dubbio importanti avanzamenti nello sviluppo di procedure pratiche e molto potenti.

3.7

Conclusioni

I modelli di fragilità e i metodi di stima dei parametri appena ricordati non sono ancora disponibili sotto forma di un *software* informatico, predefinito e completo. Tuttavia, esistono molti *software*, che richiedono alcune opzioni pratiche per l'adattamento di altri tipi di modelli di sopravvivenza o di durata. In particolare, il metodo di regressione di Cox (verosimiglianza parziale) può essere applicato fin da ora ai dati empirici. Perché rompersi il capo con i modelli di fragilità quando si può facilmente utilizzare una regressione di Cox per stimare i coefficienti delle covariate nei modelli di rischio? Ci sono tre ragioni fondamentali per fornire una risposta convincente.

Prima di tutto, il modello di regressione di Cox richiede la stima di coefficienti che tendono ad essere errati verso zero. Del resto, gli epidemiologi hanno fatto spesso rilevare che la maggior parte dei fattori di rischio che aumentano la probabilità di morte sembrano avere effetti via via meno importanti quando l'età (o la durata) aumenta. La ragione risiede nel fatto che, generalmente, gli individui ad alto rischio che sopravvivono hanno spesso energie o vantaggi nascosti, mentre numerosi individui apparentemente a basso rischio sopravvivono, ma possono essere relativamente fragili o avere dei problemi di salute in campi inesplorati. Di conseguenza, alle età più elevate o alle durate più lunghe, il gruppo a più alto rischio differisce nella sua composizione dal gruppo a basso rischio poiché comporta una minor fragilità non osservata. Se la fragilità non osservata non è inclusa nel modello, allora questo effetto condurrà ad una convergenza con l'età delle funzioni di rischio dei due gruppi (Vaupel, Manton, Stallard, 1979; Vaupel, Yashin, 1985). L'ipotesi dei rischi proporzionali, sulla quale poggia la regressione di Cox non consente di tener conto di questa convergenza: il rischio relativo stimato è una misura del rischio relativo medio dell'insieme nell'intervallo d'età. Ne risulta che

la regressione di Cox tende a sottostimare i fattori di rischio: le stime sono errate verso zero. Più in generale, ogni metodo che ignora l'eterogeneità non osservata tende a sottostimare i fattori di rischio alle età più avanzate o alle lunghe durate di esposizione.

In secondo luogo, i modelli di fragilità consentono di stimare i rischi non osservati (o rischi di base), i rischi cioè che determinano le traiettorie dei rischi individuali. Può essere utile, ad esempio, sapere se questo rischio di base aumenta uniformemente o no, anche se il rischio osservato per la popolazione aumenta all'inizio e declina in seguito. Più in generale, i demografi si chiedono se le traiettorie dei tassi per età o per durata di esposizione osservate comportano una spiegazione di livello-1 o di livello-2. La traiettoria osservata per la popolazione è ancora valida per i diversi individui che la compongono (oppure questa traiettoria è determinata da un'evoluzione della struttura)? La finalità dei modelli di fragilità è proprio quella di rispondere a questa domanda.

Infine, i modelli di fragilità consentono di utilizzare per l'analisi le statistiche di stato civile. Ad esempio, com'è stato brevemente ricordato, è possibile analizzare i dati dettagliati per un sotto-insieme di popolazione (ad esempio, i gemelli o le persone che partecipano ad un'indagine) unitamente ai dati di mortalità esaustivi di stato civile. La combinazione dei dati dettagliati per una sotto-popolazione e dei dati per la popolazione complessiva consente di giungere a stime statistiche più precise. Per l'analisi demografica questa combinazione sembra naturale e appropriata.

Riferimenti bibliografici

- ANDERSEN P. K., BORGAN Ø., GILL R., KEIDING N. (1993), *Statistical Models Based on Counting Processes*, Springer Verlag, New York.
- BLUMEN I., KOGAN M., MC CARTHY P. J. (1955), *The Industrial Mobility of Labor as a Probability Process*, Cornell University Press, Ithaca (New York) (Cornell Studies of Industrial Labor Relations, vol. 6).
- BOX G. E. P. (1976), *Science and Statistics*, in "Journal of the American Statistical Association", vol. 71, n. 356, pp. 791-802.
- COX D. R., OAKES D. (1984), *Analysis of Survival Data*, Chapman & Hall, London.
- GAVRILOV L. A., GAVRILOVA N. S. (1991), *The Biology of Life Span*, Harwood, New York.
- HALLEY E. (1693), *Mathematical Demography*, Heidelberg, Berlin-Springer, New York (ristampato in Smit, Keyfite, 1977).
- HECKMAN J. J., SINGER B. (1984), *Econometric Duration Analysis*, in "Journal of Econometrics", vol. 24, pp. 63-132.

- HORIUCHI S., WILMOTH J. R. (1998), *Deceleration in the Age Pattern of Mortality at Older Ages*, in "Demography", vol. 35, pp. 391-412.
- JEUNE B., VAUPEL J. W. (eds.) (1999), *Validation of Exceptional Longevity*, Odense University Press, Odense (Denemark).
- LE BRAS HERVÉ (1976), *Lois de Mortalite et Age Limite*, in "Population", vol. 31, pp. 655-92.
- MANTON K. G., STALLARD E., VAUPEL J. W. (1986), *Alternative Models for the Heterogeneity of Mortality Risks among the Aged*, in "Journal of the American Statistical Association", vol. 81, pp. 635-44.
- THATCHER R., KANNISTO V., VAUPEL J. (1998), *The Force of Mortality at Ages 80 to 120*, Odense University Press, Odense (Coll. Odense Monographs on Population Aging, vol. 5).
- VAUPEL J. W. (1991a), *Kindred Lifetimes: Frailty Models in Population Genetics*, in J. Adams et al. (eds.), *Convergent Questions in Genetics and Demography*, Oxford University Press, London.
- ID. (1991b), *Relatives' Risks*, in "Theoretical Population Biology", vol. 37, p. 220-34.
- VAUPEL J. W. et al. (1998), *Biodemographic Trajectories of Longevity*, in "Science", n. 280, pp. 855-60.
- VAUPEL J. W., CAREY J. R. (1993), *Compositional Interpretations of Medfly Mortality*, in "Science", n. 260, pp. 1666-7.
- VAUPEL J. W., MANTON K. G., STALLARD E. (1979), *The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality*, in "Demography", vol. 16, pp. 439-54.
- VAUPEL J. W., YASHIN A. I. (1985), *Heterogeneity's Rules: Some Surprising Effects of Selection on Population Dynamics*, in "American Statistician", vol. 39, pp. 176-85.
- VAUPEL J. W., YASHIN A. I., MANTON K. G. (1988), *Debilitation's Aftermath*, in "Mathematical Population Studies", n. 1, pp. 21-48.
- YASHIN A. I., IACHINE I. A. (1997), *How Frailty Models Can Be Used for Evaluating Longevity Limits*, in "Demography", vol. 34, pp. 31-48.
- YASHIN A. I., VAUPEL J. W., IACHINE I. A. (1994), *A Duality in Aging: The Equivalence of Mortality Models Based on Radically Different Concepts*, in "Mechanisms of Ageing and Development", vol. 74, pp. 1-14.
- ID. (1995), *Correlated Individual Frailty*, in "Mathematical Population Studies", n. 5, pp. 1-15.